Improving the Potency of Auristatin Cancer Drugs by Virtual Structure Modifications

Master's thesis
Iris Sokka

Master's Programme in Chemistry
and Molecular Sciences
University of Helsinki
November 2019

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree programme
Faculty of Science	Master's Programme in Chemistry and Molecular Science

Tekijä – Författare – Author

Iris Sokka

Työn nimi – Arbetets titel – Title

Improving the Potency of Auristatin Cancer Drugs by Virtual Structure Modifications

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	November 2019	56

Tiivistelmä – Referat – Abstract

Cancer is a worldwide health problem; in 2018 9.6 million people died of cancer, meaning that about 1 in 6 deaths was caused by it. The challenge with cancer drug therapy has been the development of cancer drugs that are effective against cancer but are not harmful to the healthy cells. One of the solutions to this has been antibody-drug conjugates (ADCs), where a cytotoxic drug is bound to an antibody. The antibody binds to specific antigen present on the surface of the cancer cell, thus working as a vessel to carry the drug specifically to the cancer cells.

Monomethyl auristatin E (MMAE) and monomethyl auristatin F (MMAF) are mitosis preventing cancer drugs. The auristatins are pentapeptides that were developed from dolastatin 10. MMAE consist of monomethyl valine (MeVal), valine (Val), dolaisoleiune (Dil), dolaproine (Dap) and norephedrine (PPA). MMAF has otherwise similar structure, but norephedrine is replaced by phenylalanine (Phe). They prevent cell division and cancer cell proliferation by binding to microtubules and are thus able to kill any kind of cell. By attaching the auristatin to an antibody that targets cancer cells, they can effectively be used in the treatment of cancer.

MMAE and MMAF exist as two conformers in solution, namely as *cis*- and *trans*-conformers. The *trans*-conformer resembles the biologically active conformer. It was recently noted that in solution 50-60 % of the MMAE and MMAF-molecules exist in the biologically inactive *cis*-conformer. The molecule changes from one conformer to the other by the rotation of an amide bond. However, this takes several hours in body temperature. As the amount of the *cis*-conformer is significant, the efficacy of the drug is decreased, and the possibility of side effects is increased. It is possible that the molecule leaves the cancer cell in its inactive form, migrates to healthy cells and tissue, and transforms to the active form there, damaging the healthy cell.

The goal of this study was to modify the structure of the auristatins so that the *cis/trans*-equilibrium would change to favor the biologically active *trans*-conformer. The modifications were done virtually, and the relative energies were computed using high-level quantum chemical methods, at density functional theory (DFT), 2nd order perturbation theory (MP2) and coupled cluster levels. Intramolecular interactions were analyzed computationally, employing symmetry-adapted perturbation theory and the non-covalent interactions analysis. The results suggest that simple halogenation of the benzene ring *para*-position is able to significantly shift the *cis/trans*-equilibrium to favor the *trans*-conformer. This is due to changes in intramolecular interactions that favor the *trans*-conformer after halogenation. For example, the NCI analysis shows that the halogen atom invokes stabilizing intramolecular interactions with the Dil amino acid; there is no such interaction between the *para*-position hydrogen and Dil in the original molecules. We also performed docking studies that show that the halogenated molecules can bind to microtubules, thus confirming that the modified structures have potential to be developed into new, more efficient and safe cancer drugs.

The most promising drug candidates are Cl-MMAF, F-MMAF, and F-MMAE where 94, 90, and 79 % of the molecule is predicted to exist in the biologically active *trans*-conformer, respectively.

Avainsanat – Nyckelord – Keywords

cancer, computational chemistry, molecular modeling, monomethyl auristatin E, MMAE, monomethyl auristatin F, MMAF, quantum chemistry, docking

Säilytyspaikka – Förvaringställe – Where deposited

E-thesis

Muita tietoja – Övriga uppgifter – Additional information

The work presented in this thesis has been partially published as a peer-reviewed article:

Sokka, I.K., Ekholm, F.S. & Johansson, M.P., 2019, Increasing the Potential of the Auristatin Cancer-Drug Family by Shifting the Conformational Equilibrium, *Molecular Pharmaceutics*, **16**, pg. 3600-3608. DOI 10.1021/acs.molpharmaceut.9b00437

Table of contents

Abbreviations

MMAE	monomethyl auristatin E
MMAF	monomethyl auristatin F
NMR	nuclear magnetic resonance
CD30	cluster of differentiation 30,[1] also known as TNFSF8, tumor necrosis factor receptor superfamily, member 8; an antigen/protein
CD79b	cluster of differentiation 79b; an antigen/protein
FDA	United States Food and Drug Administration
EU	European Union
EMA	European Medicines Agency
MHLW	Japanese Ministry for Health, Labour and Welfare
ADC	antibody-drug conjugate
GTP	guanosine triphosphate
GDP	guanosine diphosphate
SAR	structure-activity relationship
TLL	tubulin-tyrosine ligase
RB3	stathmin-like protein B3
PDB	Protein Data Bank
QM	quantum chemistry
HF	Hartree-Fock
MP	Møller-Plesset
CC	coupled cluster
DFT	density functional theory
SCF	self-consisted field
SCS-MP2	spin-component scaled second order Møller-Plesset
CCSD(T)	coupled cluster singles and doubles with perturbative triples correction
DLPNO	domain based local pair natural orbital
LSDA	local spin density approximation
GGA	generalized gradient approximation
TPSSh	hybrid Tao−Perdew−Staroverov−Scuseria density functional
DFT-D3(BJ)	3rd generation dispersion correction to DFT with Becke-Johnson damping
SAPT	symmetry adapted perturbation theory
GTO	Gaussian-type orbital
NCI	non-covalent interaction
RMSD	root-mean-square deviation
def2-QZVPPD	2nd default doubly polarized quadruple-zeta valence Karlsruhe basis set augmented with diffuse functions
def2-TZVPPD	2nd default doubly polarized triple-zeta valence Karlsruhe basis set augmented with diffuse functions
RESP	restrained electrostatic potential
VMD	Visual Molecular Dynamics

1 Introduction

In 2018 9.6 million people died of cancer worldwide, meaning that about 1 in 6 deaths was due to cancer.[2, 3] In 2016 around 34 000 new cancer diagnoses were done in Finland[4] and during their lifetime almost one third of Finns will get cancer.[5] Cancer is not just one disease, but a group of diseases.[6] When a cell undergoes DNA-damage or mutation that greatly alters its function, the cell would normally die by apoptosis (controlled cell death) induced by internal signals and signals from neighboring cells. In some cases, the cell changes in a way that it is able to avoid these safety measures and it starts to proliferate rapidly, causing cancer. The main treatment methods for cancer include surgery, radiotherapy and chemotherapy or a combination of these.[7]

Cancer drugs are usually divided into cell cycle specific and nonspecific drugs.[8] Cancer cells proliferate faster than regular cells, meaning that they go through the different cell cycle phases faster than normal cells.[6] This allows for drug treatment strategies that target certain cell cycle phases.[8] For example, for the cell to undergo cell division, the DNA needs to be doubled. This process can be targeted with antimetabolites that disturb DNA synthesis. Antimitotic agents, drugs that prevent cell division, are another example of phase specific cancer drugs.

Since many cancer drugs prevent cell division, their side-effects are due to their effects on the normal, quickly dividing cell-types of the body.[8] Such cell-types are for example cells of the bone marrow, hair follicle cells, and cells of the mucous membrane (intestinal epithelium). Therefore, common side-effects of cancer drugs include neutropenia, thrombocytopenia, hair loss, ulceration of the oral mucosa, and diarrhea. Opportunistic infections are also possible since the patient's immune system is weakened due to the bone marrow effects. Cancer drugs also have specific side-effects, for example microtubule binding antimitotic agents often cause peripheral neuropathy.[9-11]

In this work we concentrate on the antimitotic drugs monomethyl auristatin E (MMAE) and monomethyl auristatin F (MMAF). They are cell toxic substances that can be used in the treatment of cancer. They bind to microtubules, long protein filaments that are a part of the cell cytoskeleton (Figure 1).[6] Microtubules are hollow tubes formed from tubulin subunits and the site where MMAE and MMAF bind is called the vinca domain.[12-14] Binding to the vinca domain disturbs the normal behavior of microtubules preventing cell division.[15]

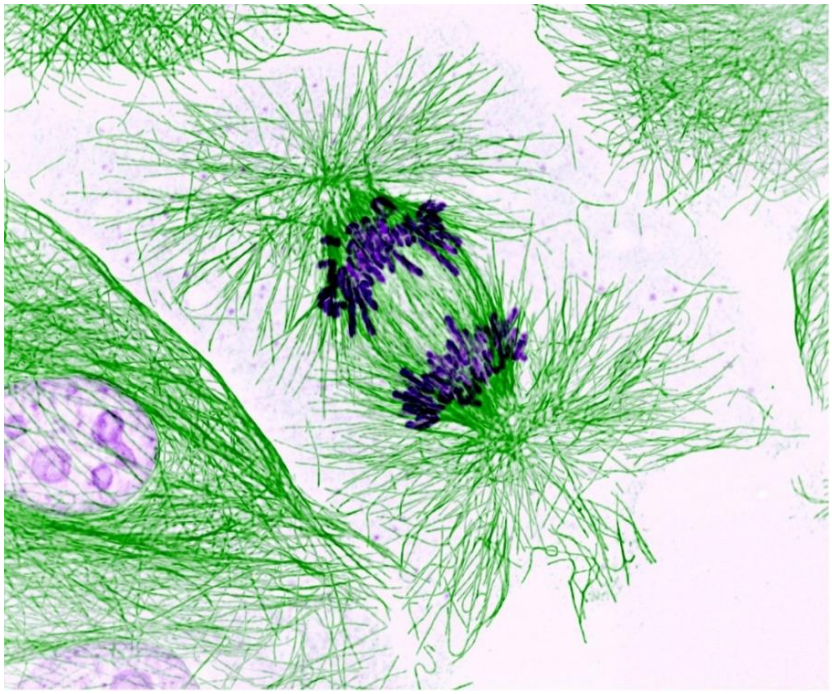

Figure 1. A cell in mitosis, the microtubules (in green) pull the chromosomes (in blue) apart to form two new cells with equal number of chromosomes. Photo by Nasser Rusan[16]

It was recently found that MMAE and MMAF exist as two distinct isomers in solution, denoted *cis* and *trans*. In their *trans*-forms the molecules bind to the microtubules preventing their normal function. This is a desired effect in cancer cells but causes adverse effects in healthy tissue. The precise three-dimensional molecular structures of the MMAE and MMAF conformers and their relative concentrations in solution was investigated by Johansson *et al.*[17] using both NMR-spectroscopy (nuclear magnetic resonance) and computational methods. According to the results, 50-60 % of the drug molecules exist in their biologically inactive *cis*-conformer in solution. The *cis*-conformer changes to the *trans*-conformer by rotation of an amide bond, but this takes several hours in body temperature. Thus, it is possible that the drug leaves the target cell in its inactive form, enters into a healthy cell and changes into the active *trans*-conformer. Inside the healthy cell the molecule can bind to its target, causing side-effects.

The goal of this study was to modify the structure of MMAE and MMAF so that the *trans*-conformer would be the more stable conformer. If the *cis*/*trans*-equilibrium could be shifted more toward the *trans*-conformer, more of the drug would be in its active form when it reaches the cancer

cell. This could potentially increase the efficiency of the drug and lead to a smaller required drug dosage. It could also improve safety of the drug, both because smaller dosage would be needed and because the drug would bind to the target when it reaches it and not migrate to healthy cells.

2 Background

Currently, there is one commercially available drug containing MMAE on the market in Europe. This drug, Adcetris, contains MMAE combined with an antibody (brentuximab vedotin). It is indicated for the treatment of adult patients with CD30-positive Hodgkin lymphoma, systemic anaplastic large cell lymphoma or CD30 positive cutaneous T-cell lymphoma.[18] CD30 is a receptor molecule whose function regulates proliferation and apoptosis of the cell.[19] A cancer where the malignant cells express CD30 protein more than regular cells of the body is called a CD30-positive cancer. Similarly, MMAE has been attached to an antibody that targets CD79b antigen (known as polatuzumab).[11, 20-22] Polatuzumab vedotin was approved by the United States Food and Drug Administration (FDA) in June 2019 for the treatment of diffuse large B-cell lymphoma in combination with other chemotherapy agents and after the disease has progressed or returned after at least two prior therapies.[23] MMAE attached to an antibody is known by the name vedotin and MMAF by the name mafodotin.[24] As of 2019[25] there are several clinical trials for vedotin/MMAE and mafodotin/MMAF with different antibodies and at different study phases.

A disease affecting less than 5 in every 10 000 people in the European Union (EU) is deemed a rare disease.[26] Over a third of all rare disease medicine are cancer medicine. For some rare diseases there is no effective treatment available. Development of new treatments is hindered by high production costs and scattered patient populations. To encourage rare disease drug development the European Medicines Agency (EMA) may grant an orphan designation to a new medicine in development. Medicines with orphan designation are offered incentives such as protocol assistance, reduced fees (for regulatory activities), access to centralized authorization procedure and 10 years of market exclusivity by the European Union.[27] Smaller pharmaceutical companies are offered additional incentives. EU works together with FDA and MHLW (Japanese Ministry for Health, Labour and Welfare) on orphan designations. Adcetris received an orphan drug designation in 2009.[28] Currently there is one MMAE containing (polatuzumab vedotin)[29] and two MMAF containing[30, 31] orphan designated medicines in development.

2.1 Cancer Drugs, Antibody-Drug Conjugates (ADC)

At the beginning of the 20th century Nobel laureate Paul Ehrlich developed the concept "Zauberkugeln" or "magic bullets".[32] These magic bullets where envisioned as drug molecules that specifically kill the target cells without causing harm to the healthy cells.[33, 34] The problem with developing cancer drugs has been the fact that cancer cells do not differ much from the healthy cells of the body. Thus, the challenge has been to produce drugs that would specifically kill cancer cells without harming normal cells. However, it has been recognized that some cancer cells express certain proteins more than the regular cells of the body.[6, 35] This has enabled the development of antibody-drug conjugates (ADC) where a cytotoxic drug is combined with an antibody that targets those proteins.[36] The operating principles and components of an ADC are shown in Figure 2. In the body the ADC migrates to the target cell and the antibody binds to the target antigen at the surface of the cell.[35, 37] The ADC is then taken into the cell by endocytosis. Inside the cell the cytotoxic agent is released from the antibody and the free drug binds to its target, inducing cell death. Thus, ADCs work like the magic bullets Ehrlich envisioned.

The properties of the warhead (cytotoxic agent), the antibody and the linker all affect the efficiency, toxicity and other properties of the ADC.[36, 38-40] The efficiency of MMAE[39] and MMAF[40] have been increased by combining them to the right antibody with a suitable linker. For example, brentuximab vedotin (Adcetris) is an ADC consisting of cell toxic MMAE attached by a linker to an antibody targeting CD30 antigen.[10, 18, 41]

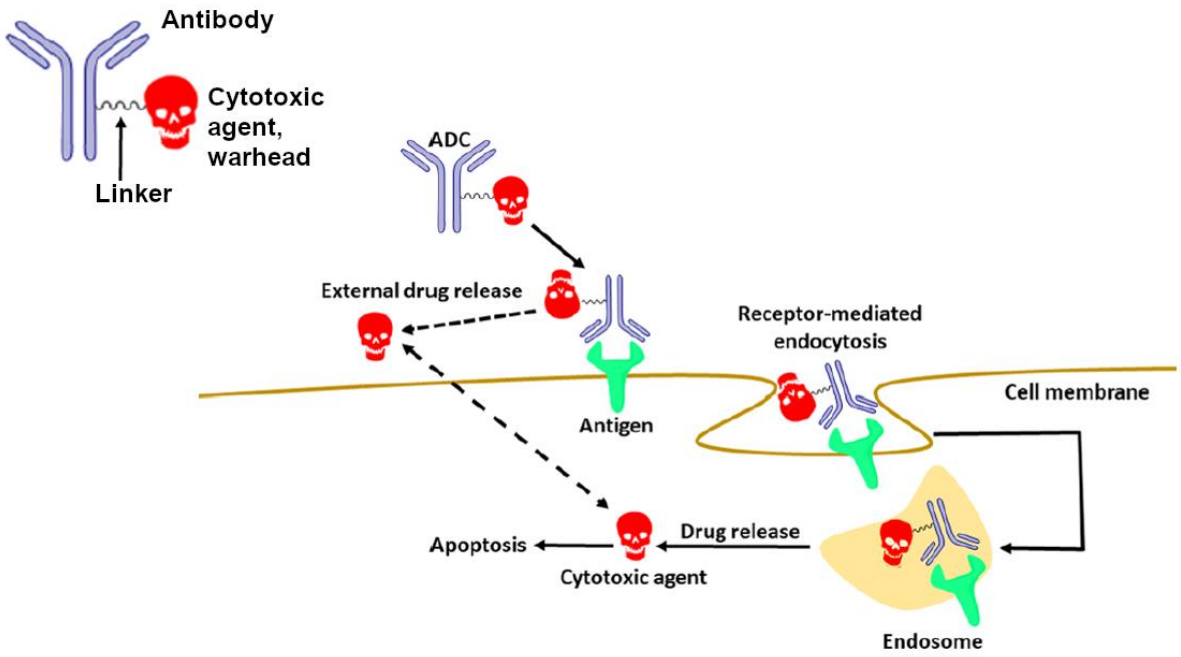

Figure 2. Structure and mechanism of an antibody-drug conjugate (ADC). Figure adapted from [17].

Microtubules are part of the cell cytoskeleton and they provide structure, enable movement, intracellular transportation, and signaling, as well as play a crucial role in cell division (mitosis) (Figure 1).[6] A microtubule is a hollow tube. Its wall is formed by αβ-tubulin dimers in rows, with 13 tubulin subunits forming the circle (Figure 3). Microtubules grow from a centrosome, a cell organelle that contains rings formed by γ-tubulin. αβ-tubulin dimers bind to γ-tubulin from the α-end (minus end) and the microtubule grows from the β-end (plus end). A guanosine triphosphate molecule (GTP) is bound to each αβ-tubulin dimer and at some time after the dimer has been added to the microtubule the GTP hydrolyses to GDP (guanosine diphosphate). The microtubule keeps growing as long as the plus end tubulin dimers have only GTP molecules (GTP cap). GTP containing tubulins are more stable and pack more tightly than the GDP containing tubulin dimers. Occasionally a GTP in the GTP cap spontaneously hydrolyses to a GDP molecule causing the microtubule to depolymerize rapidly. This behavior is known as the dynamic instability of microtubules. The microtubule becomes stable when it binds to a cell organelle.

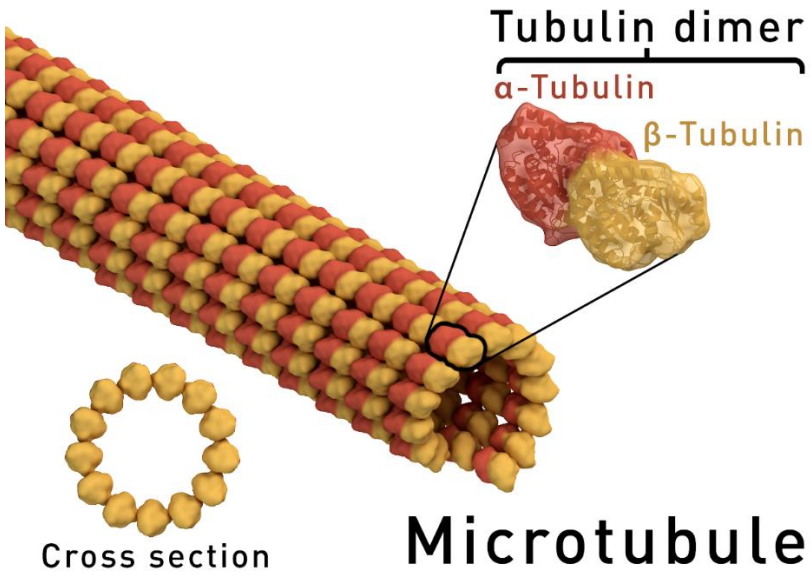

Figure 3. Microtubule is a hollow tube that is formed form αβ-tubulin dimers. The cross section shows 13 tubulin units in a ring.[42]

In cell division DNA-strands condensate into chromosomes that line up at the center of the cell (metaphase). The microtubules form nuclear spindle that pulls the chromosomes apart, splitting them into two when the cell divides (anaphase). Drugs binding to the vinca domain prevent cell division by binding to the nuclear spindle (microtubules) and stopping their normal functioning.[15] Thus, the cell is forced to remain in the metaphase, never completing the cell division; this eventually induces apoptosis.

The drugs of this study, MMAE and MMAF, belong to the vinca domain binding mitosis-inhibiting family of molecules.[12, 13] The vinca domain is located between αβ-tubulin dimers so that a vinca domain binding molecule has interactions with both the β1- and the α2-subunits (Figure 4). The auristatins bind to the vinca domain from the N-terminal end of the structure while the rest of the long structure fits to the inter-dimer interface and the hydrophobic pockets.[12] MMAE and MMAF are structurally very similar, but MMAFs higher affinity to tubulin is most likely due to the extra interaction of MMAF carboxylic acid group with β1-tubulin.[12]

When a molecule binds to the vinca domain it changes the lateral interaction between the tubulin dimers.[43, 44] When MMAE binds to tubulin, the tubulin subunits bend and twist slightly.[13] The longitudinal interaction is weaker and the position of the tubulin dimers with each other is different

than when microtubules polymerize in the absence of the drug. Thus, the microtubule does not polymerize to a straight structure, but to a curved one instead[43, 44] and the normal functioning of the microtubule is prevented. At high concentrations vinca domain binding drugs cause microtubule depolymerization, since the interaction surface between the tubulin dimers is smaller still.[15, 44] The vinca domain binding molecules used as medicine prevent mitosis in small, nanomolar concentrations that do not induce depolymerization[15] and the drug effect is due to prevention of mitosis.

Molecules that bind to the vinca domain prevent the normal functions of microtubules also by preventing the nucleotide exchange[12, 14, 45] and hydrolysis[14, 43, 45]. GTP plays an important role in the dynamic instability of microtubules. After microtubule depolymerization the GDP in a tubulin dimer has to change into GTP in order for an αβ-tubulin dimer to bind to a new microtubule again.[6] The C-terminus of tubulin-bound MMAE and MMAF is located at the vicinity of the GTP (Figure 4), preventing nucleotide exchange.[12]

2.2 Modifying the Auristatins

Both monomethyl auristatin E (MMAE) and monomethyl auristatin F (MMAF) have been developed from dolastatin 10, a cytotoxic pentapeptide that was first isolated from the sea hare *Dolabella auricularia* in 1987.[46] Dolastatin 10 binds to microtubules at the vinca domain (Figure 4) and prevents mitosis by preventing the normal function of microtubules.[14, 45] Dolastatin 10 consists of five amino acids; dolavaline (Dov), valine (Val), dolaisoleuine (Dil), dolaproine (Dap) and dolaphenine (Doe). Dolastatin 10 prevents mitosis in vitro and advanced to clinical trials, but no successful results have been reported.[47] Thus it was apparent that the structure of dolastatin 10 needed to be modified.

Structure-activity relationship (SAR) studies for dolastatin 10 by Bai *et. al.*[48, 49] revealed that configuration changes should not be made at the first three amino acids and that the new analogues should focus on finding alternative amino acid residues to replace Dap and/or Doe. Dolastatin 10 was exceptionally cytotoxic in vitro, so it was suggested that even if the modifications reduce the cytotoxicity, they should still be considered, since dolastatin 10 might be too toxic for use in patients. A less toxic, but more easily synthesized alternative for Doe should also be considered

for further development, since preparation of dolaphenine was the most difficult and expensive step of the synthesis. This led to the development of Auristatin E, where Doe has been replaced with norephedrine.[50]

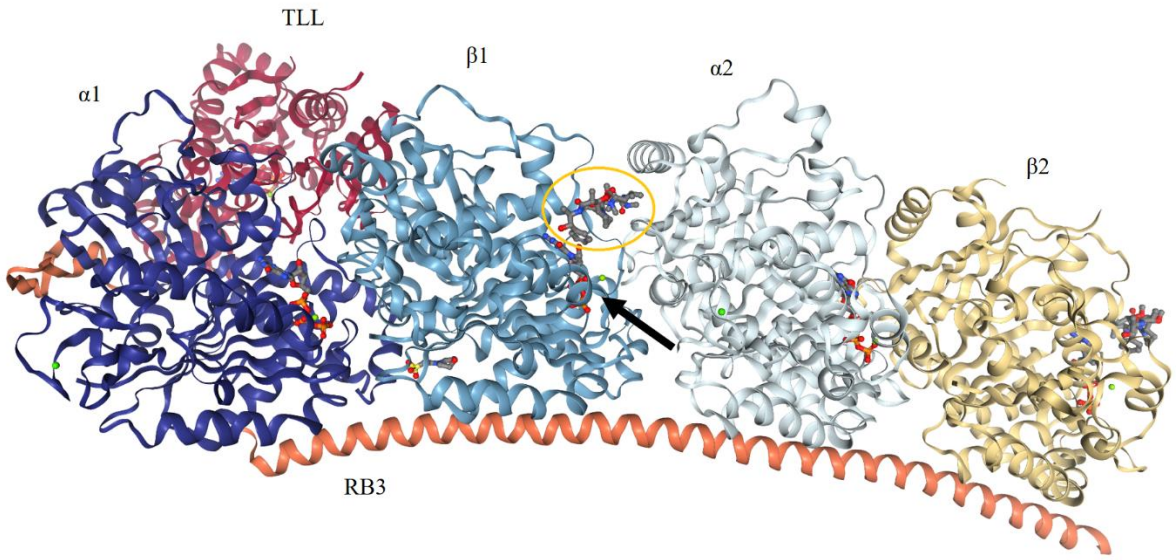

Figure 4. MMAE binding site is located at the interface of two αβ-tubulin dimers. MMAE has been circled with orange, the arrow points to GTP. PDB ID: 5IYZ[12]

Miyazaki *et. al.*[51] made several structural analogs of dolastatin 10, performing modifications in all the five amino acids to see how they affected activity (SAR presented in Figure 5). Removal of the second amino acid, valine, lead to loss of activity, meaning that all five amino acids were needed for activity. They also noted that removal of one methyl group from N-terminus did not change the activity. One of the analogs had higher antitumor activity than dolastatin 10, and it was developed further and later named soblidotin (TZT-1027).[52, 53] In soblidotin the dolaphenine moiety thiazole ring is replaced by a hydrogen. It advanced to clinical trials, but with no successful results.[47]

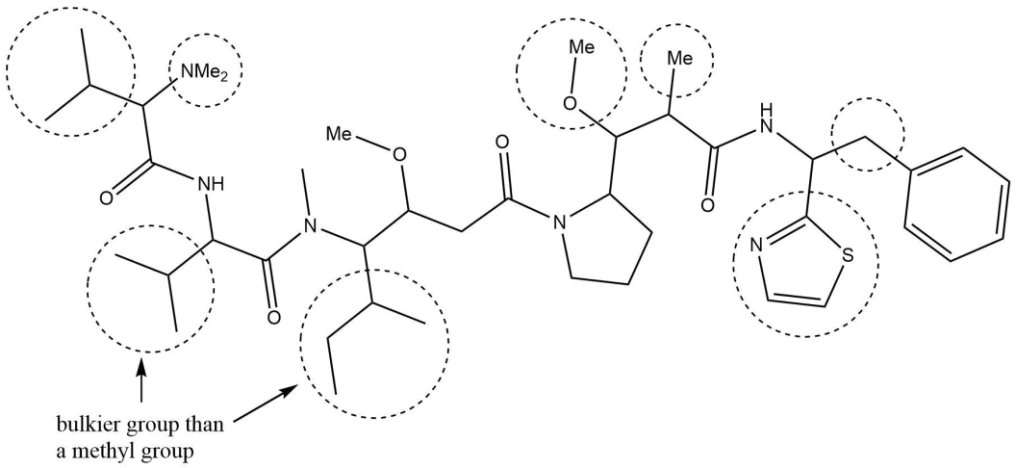

Figure 5. Dolastatin 10 SAR adapted from Miyazaki *et al*.[51] Places where structural changes have been made are circled.

Auristatin E was further developed by Doronina *et al*.[39] to be used as and ADC. For the drug to be used as a warhead molecule, it needs to be attached to an antibody. For this, they removed one of the methyl groups at the N-terminus end Dov residue, forming monomethyl auristatin E.[39, 54, 55] With ADCs in mind, monomethyl auristatin F was developed a few years later with the goal of developing a molecule that would be even more active once released to the target cell.[40]

Thus, like dolastatin 10, MMAE and MMAF are composed of five different amino acids (Figure 6). MMAE consists of norephedrine, dolaproine, dolaisoleuine, valine and monomethyl valine. MMAF has otherwise similar structure, but the norephedrine moiety is replaced by phenylalanine. In other words, MMAE has a hydroxyl group whereas MMAF has a carboxylic acid group. Due to this structural difference MMAE is a neutral molecule in physiological pH while MMAF is a negatively charged species.[12, 40, 56] Doronina *et al.*[40] also synthesized a neutral MMAF derivative, MMAF-OMe, which was highly toxic, since as a neutral molecule it can enter cells more easily than MMAF. However, instead of MMAF-OMe, MMAF has been developed into an ADC since it is effective once it reaches the target, but less toxic as a free agent.

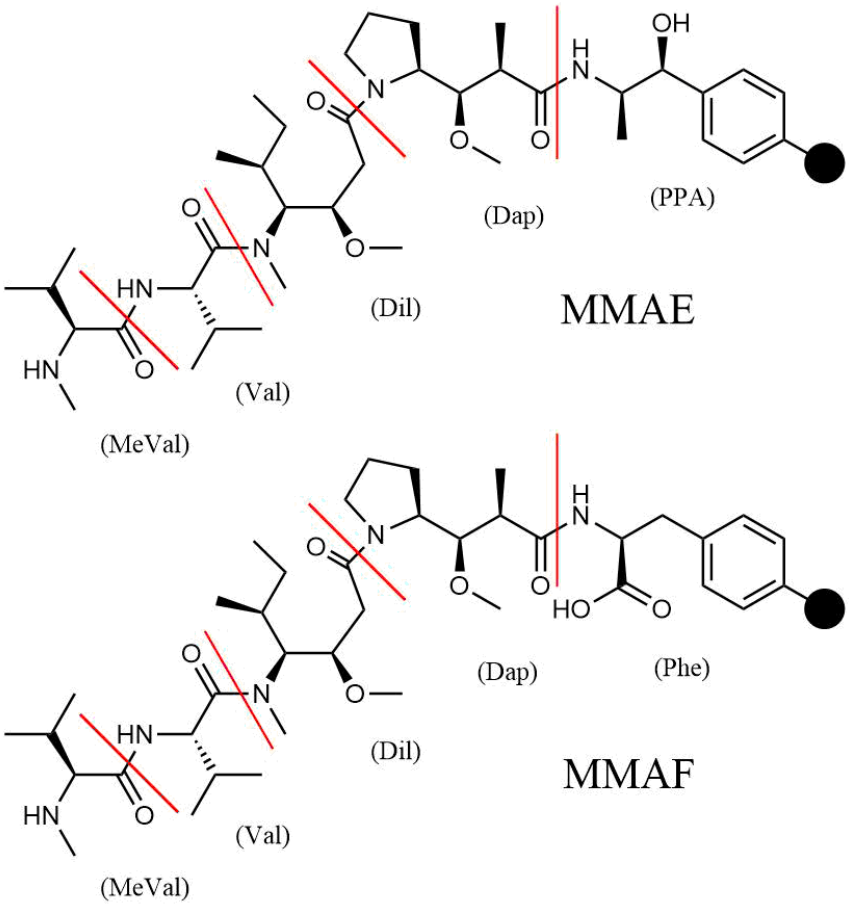

Figure 6. MMAE (above) consist of monomethyl valine (MeVal), valine (Val), dolaisoleiune (Dil), dolaproine (Dap) and norephedrine (PPA). MMAF (below) has otherwise similar structure, but norephedrine is replaced by phenylalanine (Phe). The black circle shows the site of modification. Both auristatins are shown in their *cis*-conformer.

Both MMAE and MMAF exist as two distinct, nearly isoenergetic conformers in solution, as rotation around the amide bond between dolaproine and dolaisoleuine is partially hindered.[17] The extended structure of the *trans*-conformer resembles the tubulin-bound form of the molecule while the *cis*-conformer is a more contorted compact structure (Figure 7).[12, 13, 17]

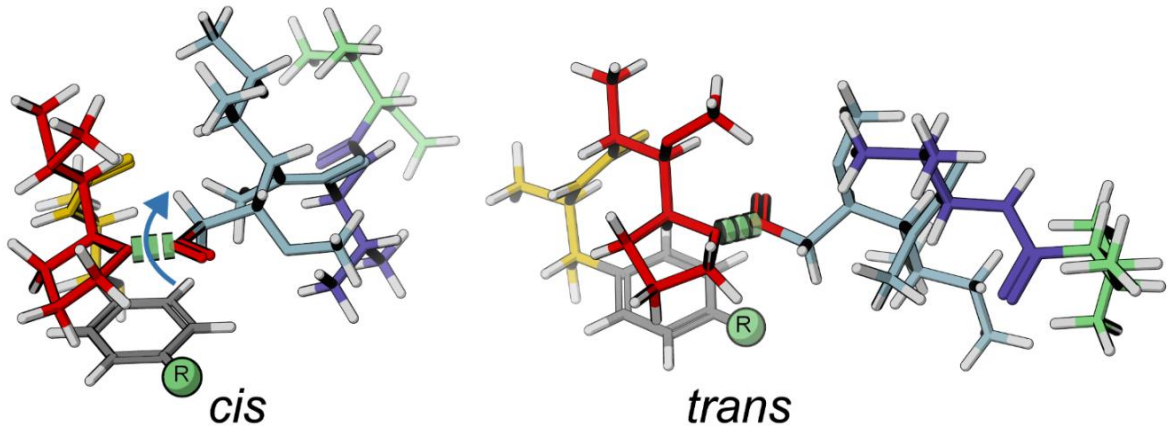

Figure 7. A 3D representation of the *cis*- and *trans*-forms of MMAE. The green discs show the peptide bond that needs to be rotate for one conformer to change to the other. The coloring shows the division for intramolecular interactions analysis, MeVal (green), Val (purple), Dil (light blue), Dap (red) and PPA (divided into yellow link part and grey phenyl part). Green ball shows the site of modification. Figure from [57].

The effects of antibody-bound and free MMAE and MMAF on certain CD30-positive cell lines *in vitro* have been studied.[40, 58] The free MMAE was found to be significantly (50- to 200-fold) more effective than free MMAF. This is because MMAF is charged and thus cannot freely pass through the cell membrane, while the neutral MMAE is able to do so. While free MMAE is more effective than free MMAF, the antibody-bound MMAF is significantly more potent than the antibody-bound MMAE.[40, 58] MMAF binds five times more strongly to tubulin than MMAE, which might explain this difference.[12] An important point to note is that neutral auristatin derivatives like MMAE and MMAF-OMe are not cell specific and would enter and kill any given cell. In order to use these cytotoxic agents as cancer drugs they need to be attached to an antibody that targets certain cells.

Antibody-bound MMAF is more potent and less toxic than similar MMAE, likely since after reaching the target cell the MMAF stays there.[39, 40, 58] Part of the potency of the MMAE containing ADC brentuximab vedotin might be explained by the so-called bystander killing effect.[10] The free MMAE exits the target cell and enters a neighboring cell causing damage. Adcetris is used in the treatment of CD30-positive Hodgkin lymphoma where the healthy cells in the microenvironment of the cancer cells support their proliferation. Destroying such neighboring cells may help kill the cancer cells faster. MMAE's ability to exit the target cell may be a benefit in treatment of such cancers, but it also increases the risk of side effects (overall toxicity).

As noted before, the three parts of an ADC, the warhead (cytotoxic agent), the antibody, and the linker, all affect the properties of the ADC.[36, 38-40] In the development of MMAE ADC Doronina *et al*.[39] tested valine-citrulline (Val-Cit) and phenylalanine-lysine linkers and noted that MMAE attached to the antibody with Val-Cit linker was more effective against the cancer cell line. Brentuximab vedotin uses the cathepsin B cleavable valine-citrulline (Val-Cit) linker.[10] The optimal drug loading for MMAE was found to be four drug molecules per antibody.[59] Eight MMAE molecules per antibody was as effective as, but not more effective than, four MMAE molecules per antibody. With eight MMAE molecules per antibody, drug clearance was higher, that is, the drug exited the body faster and the tumor exposure to MMAE was less compared to drug loading of four. With only two MMAEs per antibody, the dose needed to be doubled to have the same effect as four. Thus, it was concluded that four MMAE molecules per antibody was the optimal drug loading.

Similarly, different linkers and drug loadings were tested for MMAF.[40] MMAF attached to an antibody with maleimidocaproyl linker (mafodotin) with four MMAF molecules per antibody was found to be effective in *in vitro* assay and has been developed further. Also, other antibodies[25], linkers[60, 61] and drug delivery systems[62] have been used and other modifications to the auristatins have been made. Some of the modifications include enhancing the lipophilicity of the auristatin[63], attachment of a phosphate[64], pyrrolidine ring modifications[65], N-terminal end modifications[66] and others[66, 67].

3 Motivation

The starting point of this study is the observation by Johansson *et al.*[17] that up to 60 % of both MMAE and MMAF exist in the biologically inactive *cis*-conformation in solution. In other words, when the drug is released to the cell, only about half of the drug molecules are in the active *trans*-conformer and able to bind to the tubulins and prevent their normal function. The *cis*-conformer can change to the *trans*-conformer (and the other way around) by rotation of an amide bond. The calculated energy barrier for rotation is quite high (~100 kJ/mol), and it is estimated that the conversion from one conformer to another takes several hours in body temperature.

In general, cancer drugs cause several unwanted side effects. A common side-effect of mitosis inhibitors is peripheral neuropathy.[8-11] This is most likely caused by their mode of action: their effect on microtubules. The signal molecules of the cell move on the surface of the microtubules. A typical nerve cell consists of dendrites, the cell body (soma) and an axon.[6, 68] The somas of peripheral nerves are located in the spine and their axons may extend over a meter in length. Most signaling molecules are produced at the soma and are transported to the axon terminal by motor proteins that run along microtubules. If the microtubules are damaged, the transportation of signal molecules is disturbed, and this is observed as abnormal sensations. Therefore, damage to microtubules is observed as the symptoms of peripheral neuropathy, which include hypoesthesia (impaired sensitivity), hyperesthesia (elevated sensitivity), paresthesia (abnormal sensations), discomfort, a burning sensation, neuropathic pain or weakness.[18] The symptoms of peripheral neuropathy associated with Adcetris are due to cumulative exposure to the drug and the symptoms usually improve or resolve after the end of treatment.[18, 41] Similarly, studies with polatuzumab vedotin imply that higher doses and longer treatment times lead to increased occurrence and more severe peripheral neuropathy.[11, 69]

It is possible that some of the side-effects caused by the auristatins arise from them leaving the target cell in their inactive forms, and then entering into a healthy cell where they isomerize into the active *trans*-conformer. Inside the cell the molecule binds to its target, causing damage to the healthy cell. With a reduced amount of the inactive *cis*-conformer the likelihood of the drug exiting the target cell in its inactive *cis*-form, diffusing into the healthy tissue and changing to the active conformer is reduced. Also, the drug would not be metabolized in its inactive form before it gets to bind to the target. A smaller drug dosage would be safer in general also because a lesser amount of the drug would be in blood circulation.

We note that the issue of the two isomers could be remedied also by modifying the rotational barrier of the bond rotation that connects the two isomers. With a significantly lower barrier, the *cis*-isomer would quickly isomerize to the active *trans*-form while still inside the cancer cell. With a sufficiently high barrier, the transformation of *cis* to *trans* would be effectively prevented. Modifying the rotational barrier of a peptide bond would, however, be highly challenging. Another approach would be to lock the molecule to the *trans*-form by, for example, addition of a heteroatom

ring that contains the amide bond in the desired conformation. However, this approach would make the molecule bulkier, which might affect its binding, and more lipophilic, which would make the already poorly soluble[17] auristatin even less soluble.

Thus, the goal of this study was to modify the structure of MMAE and MMAF so that the biologically active *trans*-conformer would become the dominant conformer. This could be achieved by either making the *cis*-conformer less stable or the *trans*-conformer more stable (or a combination of the two approaches). If the *cis*/*trans*-equilibrium was shifted more completely toward the *trans*-conformer, more of the drug would be in the active form when reaching the cancer cell. This would increase the potency of the drug, meaning that a lower dose of the auristatin would be needed. Also, the safety of the drug could be improved, both directly due to a lower drug dosage and indirectly since the risk of the inactive conformer leaving the target cell and entering a healthy cell would be decreased.

4 Computational Methods

The aim of drug discovery is to find new active drug molecules. Computational methods are applied at several different stages of a drug discovery process.[70, 71] For example, they can be used to screen large libraries of molecules to identify hits (molecules that bind to the target)[70-72] as well as in designing completely new compounds or in optimizing the properties of the compounds.[70] Screening can be done experimentally, but it's slow, laborious and costly and the amount of active molecules found per molecules tested (the hit rate) is quite low.[72] In virtual screening (VS), computational methods are used to identify active molecules. VS usually has better hit rate, is faster and less costly. Several different VS strategies exist. Here we will concentrate on small molecule docking since it was used in this work.

Docking is a process where a small molecule (ligand) is placed to a binding pocket of a macromolecule (a protein, often called a receptor) and the interactions between the small molecule and the macromolecule are estimated using a scoring function.[71, 72] The binding pocket can either be pre-determined or located during the docking process. Pre-determined binding pockets are usually based on X-ray crystal structures of a known ligand to the protein. In general docking is divided into two parts, conformational search (binding pose generation) and scoring part (binding

energy evaluation). Binding poses can be predicted relatively accurately but scoring and ranking them remains a challenge.[73] Several different docking programs exist, with different strengths and weaknesses. It can be concluded that the choice of the best suited docking program depends on the application.[71, 73, 74]

One of the challenges for docking pose prediction is the choice of the protein model, as crystal structures often show only one of the many conformations the protein may have.[73] Dynamics are an important part of ligand binding in reality, as ligand binding itself may impose conformational change.[70, 72] Some docking programs are able to take protein and/or ligand flexibility into account, while some keep both the ligand and the protein rigid.

Ligand-protein interactions are evaluated by a scoring function.[70, 71, 75] Based on the score, the ligands are ranked in order of predicted biological activity. Scoring functions are usually divided to three different categories; force-field, empirical or knowledge-based scoring functions.[71, 72] Fast evaluation of binding energy requires some approximations from the scoring function. For example, many scoring functions ignore entropic changes, and solvation effects are often either crudely approximated or ignored.[71] More accurate quantum chemical methods have been used in several different applications to increase the accuracy of the binging energy prediction.[76] The main problem of using more accurate, and therefore computationally more demanding methods, is that the cost of the calculations limits both the amount of conformational sampling that can be performed as well as the number of candidate molecules to be tested.[74, 76]

In this work, computational methods that are based on quantum mechanics and quantum chemistry were used to calculate energies of the virtually modified auristatin structures. Since we aimed to improve the *cis/trans*-conformational equilibrium the focus was in the relative stability of the conformers for each molecule. We also used docking to analyze the change in protein-ligand binding the modification might cause.

4.1 Theory of Computational Methods

Quantum chemical (QM) computational methods aim to work out the energy and other properties of a molecular system by finding a solution to the Schrödinger equation. In principle, if we

disregard relativistic effects, the exact solution to the equation (Equation (1)) contains all the physical information about the system.[77]

$$\hat{H}\Psi = \mathrm{E}\Psi\,, \tag{1}$$

$\hat{H}$ is the Hamiltonian operator, Ψ is the wave function of the system and E is the energy. The Hamiltonian operator is composed of terms (operators) for the kinetic and potential energy. The kinetic energy term consists of kinetic energy of nuclei and electrons, while the potential energy term contains electron-electron, nucleus-nucleus and nucleus-electron interaction terms. The Schrödinger equation is usually solved to find a wave function that gives the lowest energy value, denoted the ground state wave-function/energy.

As the Schrödinger equation cannot be solved exactly (for systems with more than two particles), it needs to be solved approximately.[77] The Hartree-Fock (HF) approximation[78, 79] is usually the starting point for the *ab initio* wave-function methods.[80] *Ab initio* is Latin and means "from the beginning"; in this context it means that the solutions are generated only with reference to physical constants and without reference to, for example, experimental data. In HF the wave function is represented by several one electron wave functions (aka orbitals), which only indirectly interact with each other. Even if a HF calculation recovers about 99% of the total energy, it is not accurate enough for chemical problems; 1% of the total energy of a molecular system is chemically a huge number. In the HF model the electron exists in an average potential field created by all the other electrons. Thus, HF does not take individual electron-electron interactions into account. Electrons repel each other, and as they move through space, they change their path to avoid each other.[77, 81] This in turn affects the movement of all the other electrons; the movements of electrons correlate. By avoiding each other the electrons minimize repulsion and the overall energy is lowered. This energy lowering, which HF ignores, is called the correlation energy.

To obtain chemically relevant results, more accurate computational methods than HF have to be used.[77] In perturbation theory, such as Møller-Plesset (MP)[82] theory, the energy consists of the HF ground state energy and a small perturbation that accounts for the electron correlation. In coupled cluster (CC)[83, 84] theory a more accurate wave function is constructed by combining the ground state HF configuration with configurations constructed by exciting electrons from the reference (HF) orbitals to virtual orbitals. These are all wave-function methods.

Another approach is density functional theory (DFT) where the energy is calculated from the electron density. Most of the energy is calculated exactly and a small exchange-correlation energy is approximated using a functional. Figure 8 shows the computational methods in relation to each other.

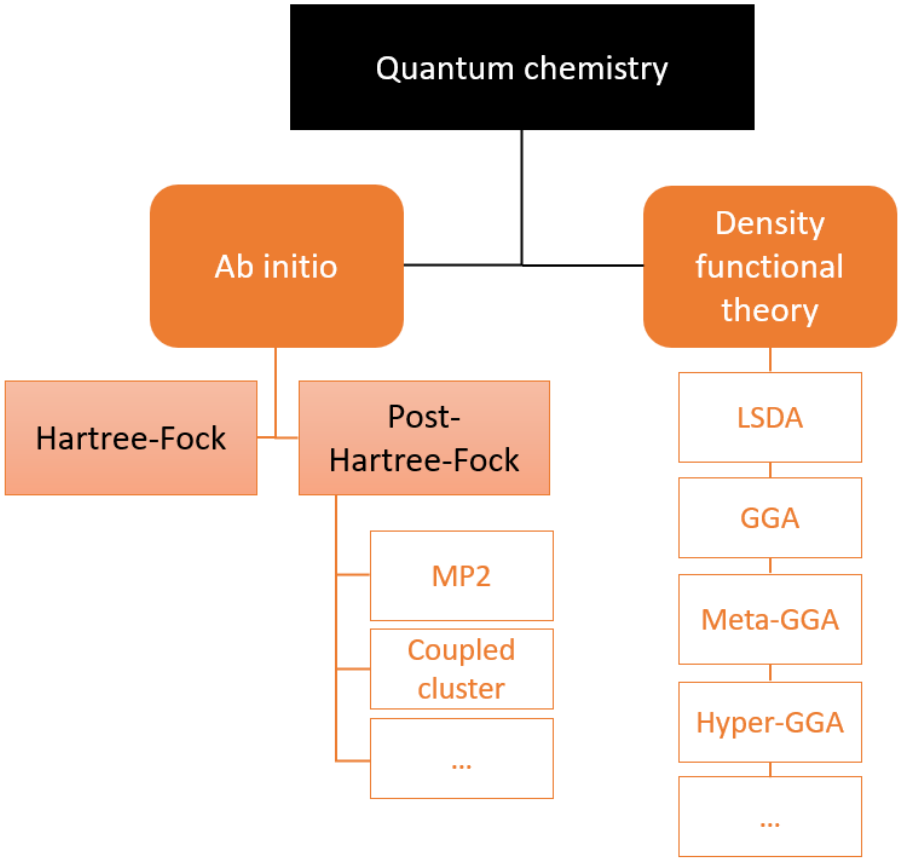

Figure 8. The computational methods used in this work are shown as a hierarchical flowchart in relation to each other. *Ab initio* methods can be divided into HF and post-HF methods (that include MP2 and CC). DFT functionals can be grouped into categories shown here and explained further in section 4.5.

4.2 Hartree-Fock Theory, HF

In Hartree-Fock[78, 79] theory the energy is calculated using a Fock operator that operates over one-electron wave functions (that consist of spatial and spin parts).[77] The one electron wave function orbitals (spatial parts) are expressed as trial functions. The Fock operator consist of the electron kinetic-energy operator, nucleus-electron attraction and average electron-electron repulsion terms. Each electron exists in an average potential field created by all the other electrons (and all the

nuclei). The calculation starts with a set of trial orbitals. These are optimized during the calculation, which is iterated until the orbitals and energy result do not improve anymore, that is, until convergence has been reached. This is called the Self-Consisted Field (SCF) method.[85]

Since the energy values are so big, even seemingly small 1% is chemically important. For example, the total energy value for *cis*-MMAE calculated at CCSD(T)-level is 6 101 793 kJ/mol, meaning that 1% of this is 61 018 kJ/mol. In comparison, the dissociation energy for a single carbon-carbon (C-C) bond is 605 kJ/mol.[86] This 1 % comes from electron correlation and taking it into account will lower the energy. The total energy for *cis*-MMAE at HF level is 6 073 115 kJ/mol, recovering 99.5 % of the total energy. HF has formal scaling of N^3.[77]

4.3 Second Order Møller-Plesset Theory, MP2

In Møller-Plesset (MP) perturbation theory, the energy consists of the ground state HF energy and an energy term that arises from perturbations (Equation (2)).[82] Second order MP (MP2)[87] is among the cheapest wave function methods that take electron correlation into account, typically returning 80-90% of the correlation energy.[77]

$$E_{MP2} = E_{HF} + \lambda E' \tag{2}$$

Electrons with parallel and antiparallel spins interact differently with each other, giving different contributions to the correlation energy. The regular MP2 theory overestimates contributions from the interaction between electrons of parallel spin, and underestimates the correlation energy between electrons of opposite spin.[88, 89] In Spin-Component Scaled MP2 (SCS-MP2) the spin contributions are scaled so that the short-ranged, opposite-spin electron correlation is enhanced, and the long-range, same-spin electron correlation energy is scaled down (Equation (3)). This simple correction usually gives significantly improved results.

$$E_{SCS-MP2} = \frac{1}{3} E_{\uparrow\uparrow} + \frac{6}{5} E_{\uparrow\downarrow} \tag{3}$$

The total energy value for *cis*-MMAE calculated at MP2-level is 6 099 009 kJ/mol. For *cis*-MMAE MP2 recovers 99.95 % of the total energy. MP2 scales as N^5.[77]

4.4 Coupled Cluster Theory, CCSD(T)

In coupled cluster (CC) theory the description of the wave function is improved by constructing it as a superposition of the ground state Hartree-Fock wave-function, and wave-functions constructed by virtually exciting electrons from the occupied HF orbitals to virtual ones.[77, 83, 84] In the CCSD method, configurations that arise from single and disconnected double electron excitations are accounted for. In the CCSD(T) method, the simultaneous excitation of three electrons are also included, but only perturbatively. This leads to scaling of N^7.[77] Thus, if the system size grows 10-fold, the calculation would take 10^7 times more time to compute. Coupled cluster theory accounts for even more electron correlation than MP2, recovering 99.99% of the total energy at CCSD(T) level.[77] Due to this high accuracy, CCSD(T) is often referred to as the gold standard of quantum chemical methods.[90, 91]

4.4.1 DLPNO-CCSD(T)

In the domain based local pair natural orbital coupled cluster method, DLPNO-CC, the computational burden of accurately computing the electron-electron interactions is reduced by two approaches: localizing the molecular orbitals and pre-screening the important energy contributions at a lower level (MP2).[92, 93]

The standard canonical orbitals extend over the whole molecule and are very diffuse. Because of this, all possible orbital pairs are close enough to each other to interact. Thus, in order to calculate the energy at the accurate CCSD(T) level, interactions between all possible orbitals would need to be calculated. In the DLPNO methods, the orbitals are first localized on certain atoms or atom groups, in other words divided into domains. Localization is possible, as the total wavefunction does not change even if the orbitals are modified by a unitary transformation. By finding an appropriate unitary transformation that creates a new set of orbitals as linear combinations of the canonical orbitals, the spatial extent of the orbitals can often be reduced. A common scheme, used also in this work, is the Foster–Boys localization method.[94, 95]

The localized orbitals are used to calculate the interaction energies at a lower level, often MP2. If the MP2 interaction energy exceeds a preset value, the energy is calculated at the more accurate

CCSD(T) level. Otherwise, the specific pair-interaction energy is approximated at the lower level of theory. Orbital localization reduces the amount of interacting orbital pairs, especially for bigger molecules, since most of the localized orbitals are spatially so far from each other that their interaction energy does not contribute significantly to the total energy. This together with a proper implementation has made DLPNO-CCSD(T) a linearly scaling method,[96] whereas regular CCSD(T) scales as N^7.[77] For small systems, the DLPNO-CCSD(T) scaling is close to standard, canonical CC, since all of the electrons are close enough to interact and thus all the interaction energies between orbitals need to be calculated accurately to obtain CC level accuracy. Large molecules, like MMAE and MMAF, which are far too big for regular CCSD(T) calculations, benefit greatly from the DLPNO scheme. Reliable coupled-cluster level results can be obtained by employing DLPNO-CCSD(T).

4.5 Density Functional Theory, DFT-D3

Hohenberg and Kohn have shown that in the absence of a magnetic field the (ground state) electron density completely and uniquely defines the energy and all other observables of a molecule[97]; integration over the electron density gives the number of electrons[98]. HF, the simplest wavefunction theory, is a 3N dimensional problem, with N the number of electrons, whereas the density depends only on the three variables of 3D space. Thus, in principle, density functional theory (DFT) is a simpler method than the wave function methods. However, in order to describe the kinetic energy of the electrons accurately, using just the total electron density is highly challenging. Therefore, orbitals are usually used. This Kohn-Sham approach makes also DFT a 3N dimensional problem.[77, 99]

In DFT most of the energy is calculated exactly and only a small exchange-correlation energy is approximated. If the form of the functional that maps electron density to energy was known, DFT calculations would give exact results. Unfortunately, this is not the case. Instead, several approximated functionals have been constructed.[100]

The energy expression contains terms for the kinetic energy (T), Coulomb repulsion between electrons (J), and nucleus-electron attraction (V_{ne}), as well as the approximate part of DFT, the exchange-correlation energy (E_{XC}) (Equation (4)).[77, 100]

$$E_{DFT} = T + V_{ne} + J + E_{XC} \quad (4)$$

The exchange-correlation term attempts to estimate the exchange and correlation energies and is often constructed as a sum of exchange and correlation energy terms.[77, 100] Several different functionals exist[100] and they can be roughly categorized on how much of the information contained in the electron density they utilize. Perdew introduced Jacob's ladder of DFT for visualizing the sequence.[81] The higher we go on Jacob's ladder, the more complex electron density derivatives are included: 1) local spin densities (local spin density approximation, LSDA) 2) gradients of the local spin densities (generalized gradient approximation, GGA) 3) orbital kinetic energy density (meta-GGA) 4) exact exchange energy density from occupied orbitals (hyper-GGA), and 5) information from unoccupied orbitals.[81, 101, 102]

In addition to the classification by Jacob's DFT ladder, density functionals can also be classified by how many ingredients from wave-function theory are included. The pure functionals have no portion of wave-function theory within their definition, while hybrid functionals usually contain a fraction of Hartree–Fock type exchange energy. Double hybrid functionals also include a MP2-like term.[103] The functional used in this work, hybrid Tao−Perdew−Staroverov−Scuseria (TPSSh)[104], can be considered to stand somewhere between the third and fourth rungs of the ladder.[77] TPSS is a non-empirical, purely meta-GGA density functional, while its hybrid form TPSSh contains 10% of Hartree-Fock-like exchange (Equation (5)).[104]

$$E_{XC}^{TPSSh} = 0.10 * E_X^{HF} + 0.90 * E_X^{TPSS} + E_C^{TPSS} \quad (5)$$

Dispersion energy is a challenge for DFT. Dispersion is a long-range phenomenon with a very small effect on the electron density. As local DFT-functionals take only the local electron density (or its gradient) as input they are not able to capture the physics of dispersion.[102, 105, 106] In Grimme's[107, 108] DFT-D3 approximation, the dispersion energy is simply added as empirical atom-pair wise specific dispersion contributions (Equation (6)). As the dispersion terms depend inversely on the distance between the atom pairs (as $1/Rij^6$), they have to be damped at short interatomic distances in order to avoid infinite attraction. In this work, we have used the Becke–Johnson type damping.[109]

$$E_{DFT-D3} = E_{DFT} + E_{D3} \quad (6)$$

4.6 Symmetry Adapted Perturbation Theory

Symmetry adapted perturbation theory (SAPT) is a quantum chemical method that can be used to calculate interaction energies.[110-112] Here we used F/I-SAPT0 (F for functional groups and I for intramolecular interaction energy between two moieties in the presence of a third moiety) to analyze the intramolecular interaction energies. For the analysis, the molecule is divided into fragments and the interaction energy between the different fragments is calculated. For MMAE the partitioning is shown in Figure 7. SAPT natively provides an energy decomposition analysis. In SAPT the energy is divided into four main parts; electrostatic, dispersion, exchange and induction energy (Equation (7)).

$$E_{SAPT} = E_{electrostatic} + E_{dispersion} + E_{exchange} + E_{induction} \quad (7)$$

The division to the four energy components is done for conceptual reasons. It should be kept in mind that only the total energy is an actual physical quantity. Thus, the division into energy components could be done in any arbitrary way. The division helps in understanding where the different interaction energy components come from, but their exact physical meaning can be debated. The electrostatic energy comes from the classical electron-electron and nucleus-nucleus repulsion and electron-nucleus attraction. The exchange energy arises from the quantum mechanical electron-electron repulsion, with origins in the Pauli principle; it can be seen as a steric repulsion term. The dispersion energy arises from momentary changes in the electron density that induces change in the charge distribution at another point in space, which leads to attraction. The induction energy comes from the interaction of the permanent multipole moments of one fragment and the induced multipole moments on the other fragment.

4.7 Basis Sets

In molecular quantum chemistry, the one-electron wave-functions (the molecular orbitals), are usually expressed as a linear combination of a certain type of mathematical functions. These functions are called basis functions, and they resemble atomic orbitals in form. Gaussian-Type Orbitals (GTOs) are often used, since they are computationally efficient. The wave function or electron density can be described by combining basis functions. The more basis functions are used, the more accurate the results are. In principle, a completely accurate calculation of the energy requires an infinite amount of basis functions.[77] However, more basis functions also lead to

computationally more expensive calculations. Thus, a compromise between accuracy and efficiency needs to be made. Different methods have different requirements for the basis set quality. It is thus important to choose an appropriate basis set for the task at hand.

Several different basis set families have been developed, and they follow a certain naming system.[77] In general, basis sets are named by how many basis functions they contain as XZ, where X is the cardinal number S, D, T, Q, etc. S stands for single (the minimum basis), D for double, T for triple etc. meaning that a DZ basis set has double the amount of basis functions compared to SZ. Since mainly valence electrons participate in chemical reactions, it is not necessary to describe the core electrons with more basis functions than minimum, and therefore most basis sets only increase the number of valence functions when increasing the basis set size. Polarization functions are important for describing bonding, as the electron density is significantly deformed from that of separate atoms. S-orbitals are polarized with p-orbitals and p-orbitals are polarized with d-orbitals and so on. For Karlsruhe-type basis sets polarization is marked as P.[113] Diffuse functions, functions that extend far from the nuclei, further improve the representation of the molecular orbitals.[114] They are important for describing, for example, anions, polarizability, and dispersion interactions properly. In this work we have used the double polarized (PP) Karlsruhe basis sets, augmented with diffuse functions (D), e.g. def-TZVPPD.

4.8 Non-Covalent Interactions Analysis

Noncovalent interactions were plotted using the NCIPLOT program.[115] It uses information contained in the electron density to calculate and color regions in space where there are non-covalent interactions between atoms.[116] Electron density (ρ) and reduced density gradient are calculated at DFT level and non-covalent interactions are recognized as low-gradient, low-density regions in space. The type of the non-covalent interaction comes from the sign of the second eigenvalue of the second derivative of the electron density ($\nabla^2\rho$) and the strength of the interaction comes from the density on the non-covalent interaction surface. Thus, dispersion interactions and hydrogen bonds can be visualized easily. An example of the results of the NCI analysis is shown in Figure 10 in section 7.2.

4.9 AutoDock

AutoDock[117, 118] is a widely used docking program: a Web of Science search from all databases with keyword "AutoDock" provides 2,573 hits by September 2019. AutoDock uses a grid-based method to rapidly evaluate binding energies of trial conformations.[118] The binding site location and size can be chosen by the user, but it is also possible to search for a binding site on the macromolecule, if a grid with larger spacing is used (although this decreases accuracy).[119] First, a grid of pairwise interaction energies between the macromolecule and ligand is calculated so that it can be used later during docking as a look-up table.[75] AutoDock incorporates some ligand and protein flexibility.[118] The scoring function used in AutoDock is force-field based and consist of terms for van der Waals interactions, hydrogen bonds, electrostatic interactions, torsional degrees of freedom, and solvation.[71, 75, 118]

During docking, the program places the ligand in the binding site, evaluates its interaction energy, goes through the conformational space by changing the torsional angles and evaluating binding energies of each conformation.[75, 119] The produced conformations are clustered together, a default root-mean-square deviation (RMSD) value for a cluster is 2 Å (for AutoDock4[118]). The clusters are ranked based on their binding energy.

5 Experiment

We aimed for a relatively simple and easy modification of auristatins that would still affect the *cis/trans*-equilibrium. Initial atomic coordinates for the MMAE and MMAF structures were taken from Johansson *et al.*[17] and the *para*-position of the benzene ring of the structures was halogenated with fluorine and chlorine. Other modifications were also tested, but none of them were equally promising. Halogenation of drug molecules has been a drug design trend.[120-122] At first it was used to improve the lipophilic properties and only later to enhance the receptor binding affinity as well as to enhance bioavailability by avoiding certain metabolic pathways.[123, 124] Changing the phenyl ring *para*-position hydrogen to a halogen is a relatively easy modification both computationally and in the synthesis of the molecule.

In practice the experiment was done by optimizing the geometries of the modified auristatins at DFT level to see the effect of the modification on the relative stability of each conformer. The

results were corroborated by calculations at MP2 and CC level. The effects of the modification on the intramolecular interaction energies were analyzed using SAPT and visualized using NCIPLOT. Since a drug needs to bind to its target in order to be effective, we performed docking studies using AutoDock to see if the modification would affect binding.

6 Methods

The original MMAE and MMAF structures were taken from Johansson *et al.*[17] The desired modification was done, and the geometry was optimized at DFT-level, using TPSSh[104] functional with Grimme's D3(BJ)[107-109] dispersion correction with Becke-Johnson damping. The doubly polarized triple-zeta def2-TZVPP basis set[113] was used, and integration performed with the fine "m4" grip.[125] Solvation effects were modeled using COSMO[126] continuum solvation model, with the dielectric constant ε set to 32.6, simulating a methanol environment, or alternatively, the dielectricity inside a cell[127, 128]. For MMAF, a structure where the carboxyl acid had turned compared to the structure used in Ref. 17 was found to be slightly lower in energy, and thus used as a starting point for the modified MMAF structures. MMAE and MMAF original molecules exist as two major conformations in solution,[17] and it was assumed that only two conformations are relevant also in the case of the halogenated auristatins.

The results at DFT-level were corroborated using wave-function based methods. Spin-Component Scaled second order Møller-Plesset theory[82, 87-89] was used in connection with the doubly polarized quadruple-zeta def2-QZVPPD basis set, augmented with diffuse functions[114]. DLPNO-CCSD(T)[92, 93, 96] was used to calculate the final electronic energies. The two-point basis-set extrapolation formula by Halkier *et al.*[129] was used to extrapolate toward the complete basis set limit from the def2-TZVPPD and def2-QZVPPD basis sets.[113, 114] The DLPNO-CCSD(T) calculations were performed with both NormalPNO and TightPNO settings[90] using def2-TZVPP and def2-QZVPP basis sets to compare the effect of the different setting. Solvation was accounted for by the COSMO[126] model, simulating an aqueous environment, with ε = 78.

Gibbs free energy at 310 K (body temperature) was estimated from harmonic vibrational frequencies, computed using analytical second derivatives.[130, 131] For the vibrational frequencies, the geometries were re-optimized in gas phase at DFT/TPSSh level, using the polarized double-

zeta/split valence def2-SVP basis set[113] and the very fine "m5" integration grid.[125] Vibrations below 50 cm^{-1} were set to 50 cm^{-1}. Intramolecular interactions were calculated using symmetry-adapted perturbation theory, F/I-SAPT0[110-112], with the augmented double-zeta jun-cc-pVDZ basis set[132] in gas phase.

Docking was conducted using AutoDock version 4.2.6.[117, 118] AutoDockTools version 1.5.6 was used to set up the calculations. The Tubulin-MMAE complex (PDB ID 5IYZ)[12] was used as the protein model. The binding site is located at the interface of chains B and C. Therefore, the other chains, waters and the original ligand (ID 4Q5) were removed before docking. The original crystal-structure ligand was used as the starting conformation, replacing the phenyl ring *para*-hydrogen with fluorine or chlorine for F-MMAE and Cl-MMAE, respectively. Standard Gasteiger charges[133] were used for the protein, and for the ligands, restrained electrostatic potential (RESP) charges[134] based on the TPSSh/def2-TZVPP electron density were computed. Ligand torsion angles were set to non-rotatable (inactive), except for the four bonds in the norephedrine moiety. AutoGrid was used to calculate a grid box of the size 40 x 60 x 40 points. The grid was centered at the original ligand position. Default values were used for the calculation. A total of 500 independent search runs, each with a maximum of 2.5 million energy evaluations, utilizing the Lamarckian genetic algorithm[75], were done for each ligand; the lowest energy docking pose was located already after 20 search runs in all cases.

The DFT and MP2 calculations were done using TURBOMOLE version 7.1-7.4[125, 135, 136], the coupled cluster calculations with Orca version 4.0.1.2[137], RESP with NWChem version 6.8[138], SAPT with PS4 version 1.1,[139] and the noncovalent interaction analysis with NCIPLOT version 3[115]. Jmol[140] and VMD[141] have been used for visualization.

7 Results

7.1 Isomer Energy Differences at Different Levels of Theory

The initial DFT calculations gave promising results, showing that halogenation is able to shift the *cis/trans*-equilibrium more toward the *trans*-conformer compared to the original auristatins. We found that the halogenation of the benzene ring creates more favorable intramolecular interactions

in the *trans*-form, stabilizing it compared to the *cis*-form, thus shifting the *cis/trans*-equilibrium toward the *trans*-form. The origins of the stabilization are discussed in the next section. Here, we begin by comparing the results of the different levels of quantum chemical theory.

In order to corroborate the DFT-based results, wave-function based methods were used, namely second order perturbation theory (section 4.3) and coupled cluster theory (section 4.4). Table 1 shows free energy difference (in kJ/mol) between the *cis*- and *trans*-conformations of the unmodified MMAE and MMAF molecules calculated with different computational methods; experimental values are included for comparison. Gibb's free energy ΔG is calculated at 310 K to model body temperature.

At DFT level the *cis/trans*-equilibrium is shifted too much toward the *cis*-conformer compared to experiment. This is either due to overestimation of the stability of the *cis*-conformer, underestimation of the stability of *trans*-conformer or a combination of both. This is an example where the computationally relatively cheap DFT method fails to be sufficiently accurate. The origin of the failure might lie in modeling of the intramolecular dispersion interactions, which have been found to be important in the case of large hydrocarbon molecules.[105, 106]

Values computed with MP2 and CCSD(T), on the other hand, agree well with the experimental results, being within 1 kJ/mol from the experimental values. These correlated wave-function methods take dispersion interactions into account natively. They are more expensive, but they also agree better with experimental values.

Table 1. Free energy difference, ΔG(310K), between the *cis*- and *trans*-conformations of the original MMAE and MMAF molecules calculated at different computational levels, as well as the experimental values. Energy values are in kJ/mol and a negative value means that the *cis*-conformer is lower in energy.

	DFT/TPSSh-D3(BJ)	**SCS-MP2**	**DLPNO-CCSD(T) NormalPNO**	**Experimental**[17]
MMAE	-7.3	-1.7	-1.3	-0.9
MMAF	-6.0	+0.1	+0.2	-0.5

Similar trends between the results at different computational levels can be seen for the halogenated structures. The energy differences (in kJ/mol) between the *cis*- and *trans*-conformers for each molecule at different levels of theory are shown in Table 2. To enable a more direct comparison of the methods, the reported energies are (pure) electronic energies without additional corrections like solvation and thermal corrections. A negative value means that the *cis*-conformer is lower in energy.

The data shows that halogenation of the molecules changes the *cis/trans*-equilibrium to favor the *trans*-conformer more than for the original molecules. At DFT level, the *cis*-conformer is favored less than before: the original MMAE *cis/trans* difference is -11.6 kJ/mol, whereas the difference is -2.9 kJ/mol for F-MMAE and -2.8 kJ/mol for Cl-MMAE. In other words, the relative energy change is 8.7-8.8 kJ/mol. In the case of MMAF the *cis/trans* energy difference goes from -10.6 kJ/mol to -1.7 kJ/mol for F-MMAF (change of 8.9 kJ/mol) and to -0.0 kJ/mol for Cl-MMAF (change of 10.6 kJ/mol).

Table 2. Comparison of gas phase electronic energies calculated with different computational methods. The energies are in kJ/mol, and the values are *cis-trans* energy difference, where a positive value means that the *trans*-conformer is lower in energy.

	DFT/TPSSh-D3(BJ)	SCS-MP2	DLPNO-CCSD(T) NormalPNO
MMAE	-11.6	-6.0	-5.6
F-MMAE	-2.9	+4.3	+4.6
Cl-MMAE	-2.8	+3.5	-2.2
MMAF	-10.6	-4.6	-4.4
F-MMAF	-1.7	+4.4	+7.0
Cl-MMAF	-0.0	+6.4	+5.2

At the more accurate MP2 level the *cis/trans*-equilibrium favors the *trans*-conformer for all halogenated species (Table 2). For the original MMAE molecule, the *cis*-conformer is favored with the *cis-trans* energy difference being -6.0 kJ/mol. Halogenation changes this to favor the *trans*-conformer with +4.3 kJ/mol for F-MMAE and +3.5 kJ/mol for Cl-MMAE (with change of 10.3 and 9.5 kJ/mol respectively). For MMAF the results are similar, the energy difference goes from

-4.6 kJ/mol for original MMAF to +4.4 kJ/mol for F-MMAF and +6.4 kJ/mol for Cl-MMAF (change of 9.0 and 11.0 kJ/mol respectively).

At coupled cluster level the *trans*-form is favored for F-MMAE, F-MMAF and Cl-MMAF (Table 2). For MMAE the energy difference goes from -5.6 kJ/mol to +4.6 kJ/mol for F-MMAE (change of 10.2 kJ/mol). For Cl-MMAE the *cis*-conformer is favored, but less than for the original molecule. The *cis/trans* energy difference is -2.2 kJ/mol for Cl-MMAE, with a change of only 3.4 kJ/mol from original to chlorinated species. For MMAF the original favors *cis*-conformer with energy difference of -4.4 kJ/mol. This changes to favor the *trans*-conformer as +7.0 kJ/mol for F-MMAF (change of 11.4 kJ/mol) and +5.2 kJ/mol for Cl-MMAF (change of 9.6 kJ/mol).

Overall the *cis/trans*-equilibrium values obtained at MP2 and CC levels agree with each other and differ from the DFT estimates. The DFT results consistently predict the *cis/trans*-equilibrium to favor the *cis*-conformer more than the other methods do. However, the change in relative *cis/trans*-equilibrium for each modified molecule in comparison to the original is about the same magnitude, regardless of the method used (9-11 kJ/mol, with the exception of Cl-MMAE at CC level). Thus, although absolute *cis/trans* energy difference at DFT level differs from the MP2 and CC results, the change in the relative energies is similar. This shows that the relative energies are not overly sensitive to the used method, most likely due to error cancellation.

Table 3 shows the final energies and *cis/trans*-equilibriums calculated with DLPNO-CCSD(T) for the original and modified molecules. The middle column values include electronic energy, free energy ($\Delta\Delta G$) and solvation corrections (water as solvent). It can be noted that the simulated relative energies for MMAE and MMAF are the same whether the thermal corrections are calculated at 295 K or 310 K.[57] The calculated energy differences for the original MMAE and MMAF agree well with the experimental values. From these final, best-estimate results it can be concluded that the *cis/trans*-equilibrium can indeed be improved by fluorination and chlorination at the *para*-position of the C-terminal phenyl group. The results for halogenated species show notable improvement in the *cis/trans*-equilibrium to favor the biologically active conformer. The change is significant for Cl-MMAF, F-MMAF, and F-MMAE, with the fraction of *trans*-conformer being 94, 90 and 79 %, respectively.

Table 3. Final *cis/trans*-equilibrium energies on DLPNO-CCSD(T) level, in kJ/mol. Solvation to water. (*cis*:*trans* ratio in parentheses)

	MMAE	MMAF	MMAE	MMAF	MMAE	MMAF
	ΔH, 310 K		ΔG, 310 K		ΔG, 295 K, exp[17]	
original	-1.0 (59:41)	+0.1 (49:51)	-1.3 (62:38)	+0.2 (48:52)	-0.9 (**59:41**)	-0.5 (**55:45**)
para F	+4.0 (17:83)	+7.0 (6:94)	+3.4 (**21:79**)	+5.6 (**10:90**)		
para Cl	+1.2 (39:61)	+7.4 (5:95)	+0.3 (47:53)	+7.2 (**6:94**)		

We end this section with a discussion on computational efficiency. Figure 9 shows the formal scaling of the wave function methods. HF scales as N^3, MP2 as N^5 and CCSD(T) as N^7. Using DLPNO-CC one can obtain CCSD(T) -level accuracy with linear scaling. Without the use of DLPNO such high accuracy results for MMAE and MMAF could not be obtained in reasonable time. DFT scales as N^3 like HF, but the DFT results are accurate enough that they are applicable to chemical problems, unlike HF results. Thus, it can be noted that the advantage of DFT is that it is a cheap method with relatively accurate results. More precise wave function methods, like MP2 and CCSD(T) used here give results that agree well with experimental values. The applicability of a certain methods depends on the problem.

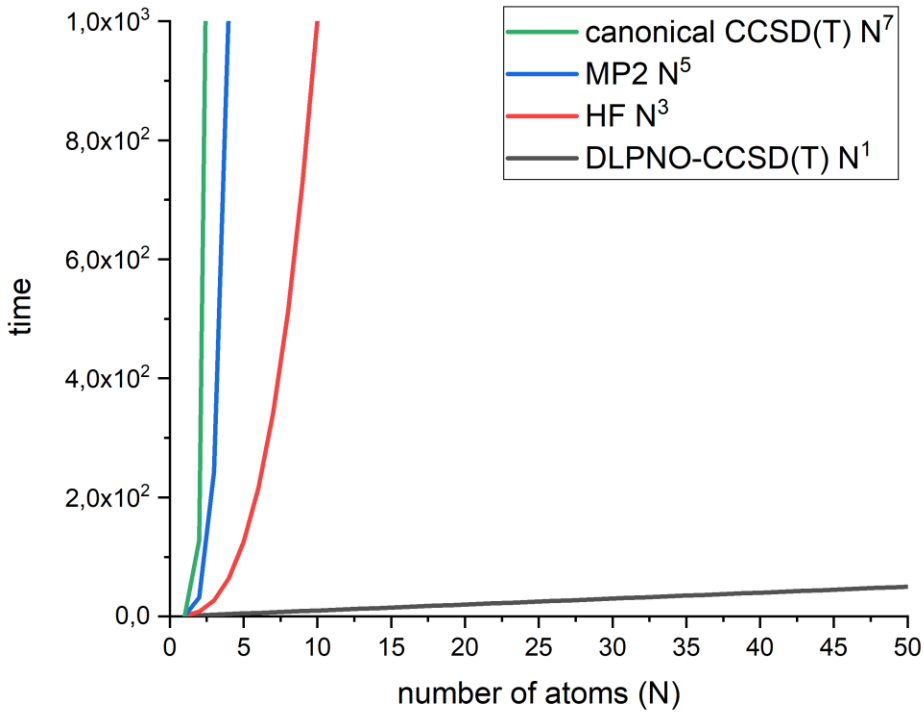

Figure 9. Formal scaling of wave function methods. Hartree-Fock (HF) scales as N^3 (red), Møller-Plesset (MP2) scales as N^5 (blue), canonical coupled cluster (CCSD(T)) scales as N^7 (green) and domain based local pair natural orbital CC (DLPNO-CCSD(T)) has linear scaling (black).

7.2 Intramolecular Interactions Energies

Table 4 shows intramolecular interactions calculated with symmetry-adapted perturbation theory, F/I-SAPT0. For each calculation the molecule was divided into three parts, the phenyl ring, a small bridge part, and the rest of the molecule. To avoid dividing the system over an amide bond, the cut was done after the carbonyl carbon of the Dap amino acid moiety. The divided parts are shown with coloring in Figure 7, where the norephedrine phenyl ring is colored in grey, the bridge part in yellow and the for the rest of the molecule the amino acids are colored with red, blue, purple and green. The intramolecular energy is divided into four parts, electrostatic, dispersion, exchange repulsion, and induction. The total interaction energy is more attractive in the *trans*-conformer after halogenation for both MMAE and F.

The halogen interacts with the Dil moiety of the molecule in the *trans*-form of the compound (Figure 10). This interaction explains the greater relative stability of the *trans*-form of the halogenated species. Electrostatic and dispersion interactions favor the *trans*-conformer more, and induction does not change much. Exchange energy difference changes to favor *cis*-conformer more than for the original; this is probably due to more repulsion between the electrons of the halogen and the rest of the structure in the *cis*-conformer. Overall the total interaction energy goes from favoring the *cis*-form by -7.4 kJ/mol for original MMAE to favoring *trans*-form by +4.4 kJ/mol for F-MMAE and +4.2 kJ/mol for Cl-MMAE. For MMAF the values are -9.0 kJ/mol for the original to +3.3 for F-MMAF and +2.8 kJ/mol for Cl-MMAF.

Table 4. Intramolecular interaction energies between the terminal phenyl ring and the four remaining peptides, based on a F/I-SAPT0 analysis. The interaction energy is divided into four parts, electrostatic, dispersion, exchange, and induction. Energy values are in kJ/mol. A positive value means that the equilibrium favors *trans*-conformer.

MMAE	**total**	**electrostatic**	**dispersion**	**exchange**	**induction**
original	**-7.4**	-7.1	-8.5	+9.8	-1.7
para F	**+4.4**	+8.1	-1.4	-2.3	+0.0
para Cl	**+4.2**	+6.4	-4.8	+3.7	-1.1
MMAF	**total**	**electrostatic**	**dispersion**	**exchange**	**induction**
original	**-9.0**	-8.3	-9.7	+10.9	-1.9
para F	**+3.3**	+4.2	-6.4	+6.6	-1.1
para Cl	**+2.8**	+4.4	-5.6	+5.1	-1.1

Next, we take a closer look at the different energy-components of the intramolecular interaction between the terminal phenyl group and the rest of the auristatin. The electrostatic interactions are more attractive in the halogenated *trans*-conformers, whereas for the original auristatins the *cis*-conformers experience greater electrostatic interactions. For MMAE the electrostatic difference between *cis/trans* goes from -7.1 kJ/mol for the original to +8.1 kJ/mol for F-MMAE (change of 15.2 kJ/mol) and +6.4 kJ/mol for Cl-MMAE (change of 13.5 kJ/mol). For MMAF the values are -8.3 kJ/mol for the original to +4.2 for fluorinated and +4.4 kJ/mol for chlorinated species (change of 12.5 and 12.7 kJ/mol respectively).

For the original molecules, dispersion favors the *cis*-conformer (-8.5 kJ/mol for MMAE and -9.7 kJ/mol for MMAF). For the halogenated species, this changes, and dispersion does not favor the *cis*-conformer as much as before: -1.4 kJ/mol for F-MMAE, -4.8 kJ/mol for Cl-MMAE, -6.4 kJ/mol for F-MMAF, and -5.6 kJ/mol for Cl-MMAF. Thus, dispersion still favors the *cis*-conformer, but not as much as for the original molecules. The induction energy difference between the *cis*- and *trans*-conformers for each molecule remains almost the same.

For the original molecules, there is relatively more exchange repulsion in the *cis*-conformer than *trans*-conformer, since in the more compact *cis*-conformer the atoms are closer together and repel each other more. Halogenation changes the small hydrogen to a bigger atom with more electrons; this creates more e-e repulsion for both of the conformers. For the relative *cis/trans* exchange energy this means that the for the original molecules the *trans*-conformer is favored more than in the halogenated species; in the case of F-MMAE, the *cis*-conformer actually experiences less steric repulsion than the *trans*-conformer.

Thus, it can be concluded that while halogenation does induce unfavorable e-e repulsion, this is more than compensated for by favorable electrostatic and dispersion interactions, so that the overall interaction energy is more attractive for the *trans*-conformer.

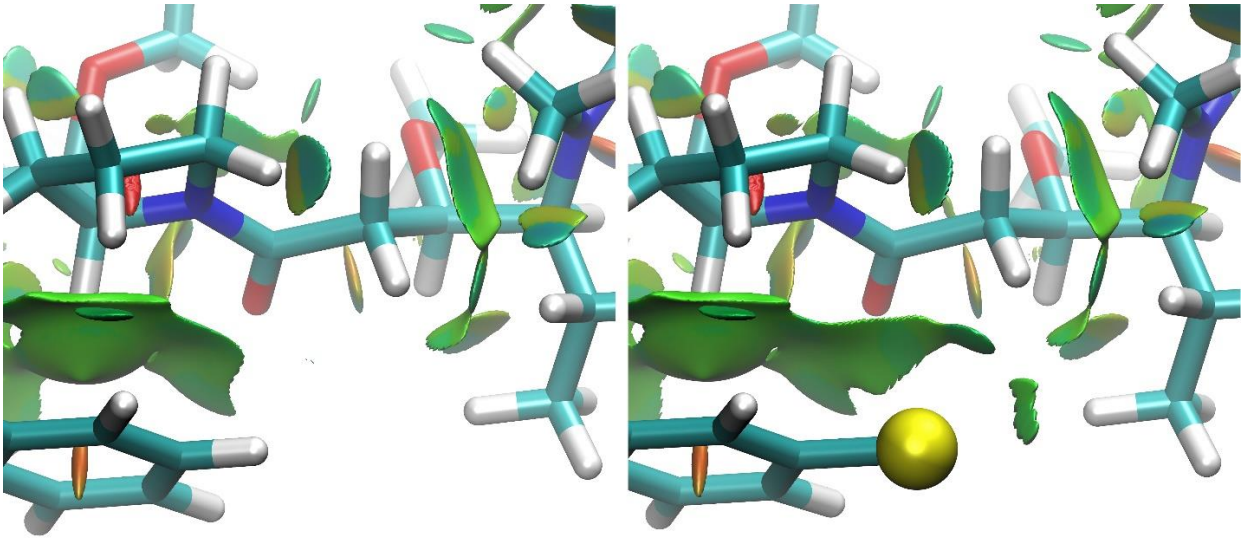

Figure 10. Non-covalent interaction plot for MMAE and Cl-MMAE *trans*-conformers. The green areas indicate intramolecular dispersion interactions. Compared to the original MMAE there is more interaction energy between the benzene ring and the rest of the structure in Cl-MMAE due to halogenation. Especially the chlorine interacts with the Dil amino acid. Figure from [57].

The noncovalent interactions can be plotted as shown in Figure 10. The green indicates that there are dispersion interactions (the theory in section 4.8). In Cl-MMAE there is more interactions between the halogenated benzene ring and the rest of the structure, compared to the original MMAE.

7.3 Docking

A docking study of MMAE, F-MMAE, and Cl-MMAE binding to microtubules was performed using the AutoDock protocol. The lowest energy binding poses for each ligand as well as the original crystal structure of the MMAE molecule at the tubulin binding site are shown in Figure 11. AutoDock places MMAE very close to the experimental crystal structure MMAE ligand. For the modified structures, the halogenated phenyl rings have moved slightly so that the halogen can fit at the binding site. The benzene rings of the fluorinated and chlorinated species turn in different directions. The lowest energy poses of both F-MMAE and Cl-MMAE have a 1.1 kJ/mol weaker binding energy than the original MMAE. The mean binding energy of lowest binding energy clusters of both F-MMAE and Cl-MMAE were also weaker, by 1.5 and 1.8 kJ/mol, respectively. This suggests that the halogenated MMAE structures bind slightly weaker than the original

molecule, but the change is small. It should also be noted that the protein structure that was used is the MMAE-tubulin crystal structure which might favor the original molecule since the protein was crystallized with the ligand at the binding pocket. Overall, the docking indicates that halogenation of the auristatins does not change binding much. Thus, it can be assumed that the modified molecules would bind to tubulin in reality as well.

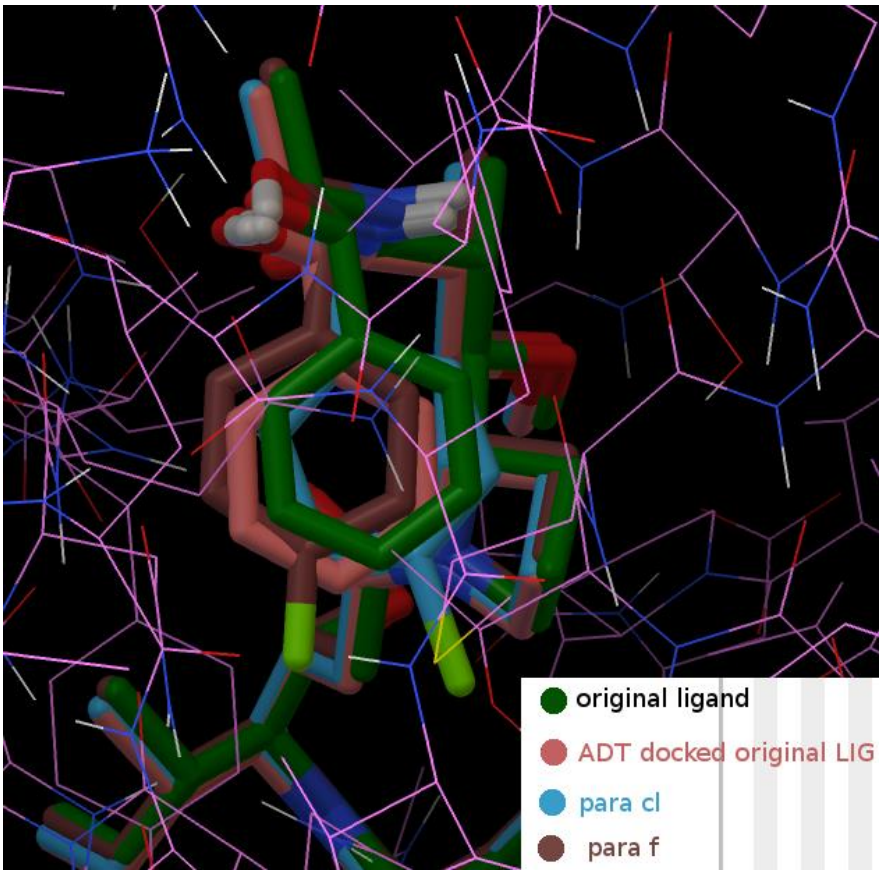

Figure 11. Docking of MMAE, F-MMAE and Cl-MMAE with AutoDock shows how the halogenated phenyl ring moves slightly to so that the halogen can fit in the binding pocket. The original crystal structure ligand MMAE is shown in dark green, MMAE docked with AutoDock is shown in terracotta, Cl-MMAE in light blue and F-MMAE in brown. The halogen atoms are colored with light green, oxygens with red, hydrogens with white and nitrogens with blue. The protein carbons are colored in pink and styled as sticks.

7.4 Error Sources

For the isomer energy differences studied here, high accuracy is crucial. Already a 1 kcal/mol (4.184 kJ/mol) energy difference leads to Boltzmann distribution of 84:16. Thus, the error in energy

needs to be small in order to obtain meaningful distribution. The focus here is in the relative energy difference between the *cis*- and *trans*-conformers. Thus, to some degree, error cancellation and systematic errors that are practically of equal magnitude for the two isomers simplifies the task of obtaining accurate final equilibrium energies. In this section we take a look at some potential sources of errors and uncertainties in the presented calculations.

The DLPNO-CC calculation may be done with different settings, Loose-, Normal- or TightPNO.[91] These setting change various thresholds that control when a full CCSD(T) calculation will be performed for a specific electron-electron interaction, and when it will be approximated at a lower level of theory. Previously, it has been demonstrated that on average, one can expect an accuracy of *ca* 4 kJ/mol and 1 kJ/mol compared to canonical CCSD(T), using the NormalPNO and TightPNO setting, respectively.[90, 91] In order to investigate how the different settings affect the results in the present case, we calculated the DLPNO-CCSD(T) *cis-trans* energy difference for MMAE with Normal- and TightPNO settings. The results are shown in Table 5.

Table 5. DLPNO-CCSD(T) calculations with different basis sets and Normal- and TightPNO settings for MMAE. The values are in kJ/mol.

MMAE	DLPNO-CCSD(T) with different basis sets	*cis* - *trans*, kJ/mol
NormalPNO	[T,Q]ZVPPD extrapolation	-5.6
	TZVPP	-7.5
	QZVPP	-7.2
	[T,Q]ZVPP extrapolation	-7.1
TightPNO	TZVPP	-9.9
	QZVPP	-9.8
	[T,Q]ZVPP extrapolation	-9.9

The TZVPP basis set has fewer basis functions than the QZVPP basis set, and thus the calculation with the QZ basis set takes electron correlation better into account since the orbitals are described more accurately. With TightPNO settings, the *cis*-conformer is favored more than for the NormalPNO settings. Comparing this to experimental values, results obtained by the looser thresholds of the NormalPNO settings deviate less from the experimental values; -7.1 kJ/mol for

NormalPNO and -9.9 kJ/mol for TightPNO. When we use diffuse functions the *cis/trans* difference agrees even better with experiment; for calculations with NormalPNO settings the energy difference is -5.6 kJ/mol using the [T,Q]ZVPPD basis set in comparison to -7.1 kJ/mol for the [T,Q]ZVPP basis set extrapolation without diffuse functions.

Here, it is appropriate to note that in our study[57] we erroneously state that the energies calculated at DLPNO-CCSD(T) level were done using the TightPNO settings, while they in fact were done with the NormalPNO settings. In any case, the excellent agreement of results calculated with NormalPNO settings with the experimental values of the original auristatins indicate that the chosen level of theory is sound. The small difference to the TightPNO results, which is in line with previous studies, does indicate that there is some amount of error cancellation involved.

The largest error sources can be estimated to be the enthalpy/entropy and solvation corrections. In this work, the Gibbs free energy was approximated from harmonic vibrational frequencies. Thus, for example, methyl group rotations are approximated as vibrations. For solvation we used a continuum solvent model instead of explicit water molecules. It can be assumed that the translational, rotational and vibrational motions contribute to the energy a different amount in gas and solution phase.[142] Thus, the solvation energy was calculated by adding the energy calculated using gas phase structures and frequencies to solvation energy calculated using COSMO.[143] In some cases, especially when the gas and solution phase structures differ significantly, using solution phase vibrational frequencies to obtain solvation energy may be more accurate than the approximations used here.[144]

Overall, the accuracy of the computational models used here is quite good. Computational methods are usually not accompanied by error bars and estimating the accuracy of a given calculation needs to be based on comparisons to either more accurate calculations or experiment. Where a point of reference is lacking, as in the case of the halogenated auristatins in this work, an educated guess is often as good as it gets. Here we have shown that the MP2 and CC level calculations agree well with experiment, as the *cis/trans*-equilibrium results are within 1 kJ/mol of experimental values. *A priori*, there is no reason to suspect that the errors for the halogenated species would be notably larger.

8 Conclusions and Outlook

We have presented a novel approach for increasing the potency of the auristatin cancer drug family; shifting the conformational equilibrium from the biologically inactive toward the active conformation. While other modifications to the auristatins have been made before, they are based on different design strategies, with a focus on e.g., lipophilicity, delivery, or affinity to tubulins.

Here, we have shown that a simple halogenation at the *para*-position of the C-terminal phenyl ring of the auristatins has the desired effect on the *cis/trans*-equilibrium. As the proposed changes in the structure are (relatively) small, the synthesis should be relatively easy. We have shown, using docking studies, that the modified molecules should bind to the tubulin in a similar manner as the original, unmodified MMAE and MMAF. The added group is quite stable and addition of fluorine and chlorine to a drug molecule is a widely used drug modification strategy.

The goal of the study was to changes the *cis/trans*-equilibrium so that more of the molecule would be in the active *trans*-conformer than for the original MMAE and MMAF. This was achieved and the same result can be seen on all different computational theory levels. In the case of DFT, the *cis/trans*-equilibrium of the modified structures still favors the *cis*-conformer, but significantly less than for the original molecules. At MP2 and CC level the equilibrium changes to favor the *trans*-conformer. Still, the change in relative energies between the originals and modified structures is of the same magnitude on all levels of theory.

As can be seen for the intramolecular interaction energy analysis (Table 4) and non-covalent interactions plot (Figure 10), the halogen creates and increases existing favorable interactions in the *trans*-conformer. The NCI plot shows that there is new interaction between the halogen and Dil amino acid.

In conclusion, halogenation of the auristatin benzene ring *para*-position changes the *cis/trans*-equilibrium to favor the *trans*-conformation. This is significant for Cl-MMAF, F-MMAF and F-MMAE as 94, 90 and 79 % of the molecule is in the *trans*-form (respectively). Next, based on this study, modified auristatins will be synthesized and the *cis/trans*-equilibrium will be measured.

Also, the biological activity will be measured in cancer-cell lines. Auristatins modified using the presented design strategy have the potential to be developed into new ADC-drugs.

9 References

1. Chan, J.K., Ng, C.S. & Hui, P.K., 1988, A simple guide to the terminology and application of leucocyte monoclonal antibodies, *Histopathology,* **12**, pg. 461-480. DOI 10.1111/j.1365-2559.1988.tb01967.x.

2. *Cancer - WHO*. Available: http://www.who.int/en/news-room/fact-sheets/detail/cancer (Retrieved Jun 20, 2019).

3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. & Jemal, A., 2018, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer Journal for Clinicians,* **68**, pg. 394-424. DOI 10.3322/caac.21492.

4. Malila, N., Pitkäniemi, J., Virtanen, A., 2019, *Syöpä 2016 -raportti*.

5. *Syövän yleisyys - Kansantaudit - THL*. Available: http://thl.fi/fi/web/kansantaudit/syopa/syovan-yleisyys (Retrieved Jun 14, 2018).

6. Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2010, *Essential Cell Biology*, 3rd ed., Garland Science, New York.

7. Katzung, B.G., 2018, *Basic & Clinical Pharmacology*, 14th ed., McGraw-Hill Education, New York; Chicago; San Francisco; Athens; London; Madrid; Mexico City; Milan; New Delhi; Singapore; Sydney; Toronto.

8. Koulu, M., Mervaala, E., Tuomisto, J., 2012, *Farmakologia ja toksikologia*, 8th ed., Kustannusosakeyhtiö Medicina, Kuopio.

9. Lee, J.J. & Swain, S.M., 2006, Peripheral neuropathy induced by microtubule-stabilizing agents, *Journal of Clinical Oncology,* **24**, pg. 1633-1642. DOI 10.1200/JCO.2005.04.0543.

10. Katz, J., Janik, J.E. & Younes, A., 2011, Brentuximab vedotin (SGN-35), *Clinical Cancer Research,* **17**, pg. 6428-6436. DOI 10.1158/1078-0432.CCR-11-0488.

11. Palanca-Wessels, M.C.A., Czuczman, M., Salles, G., Assouline, S., Sehn, L.H., Flinn, I., Patel, M.R., Sangha, R., Hagenbeek, A., Advani, R., Tilly, H., Casasnovas, O., Press, O.W., Yalamanchili, S., Kahn, R., Dere, R.C., Lu, D., Jones, S., Jones, C., Chu, Y.-W. & Morschhauser, F., 2015, Safety and activity of the anti-CD79B antibody-drug conjugate polatuzumab vedotin in relapsed or refractory B-cell non-Hodgkin lymphoma and chronic

lymphocytic leukaemia: A phase 1 study, *The Lancet Oncology*, **16**, pg. 704-715. DOI 10.1016/S1470-2045(15)70128-2.

12. Waight, A.B., Bargsten, K., Doronina, S., Steinmetz, M.O., Sussman, D. & Prota, A.E., 2016, Structural basis of microtubule destabilization by potent auristatin anti-mitotics. *PLoS One*, **11**, pg. e0160890/-e0160890/14. DOI 10.1371/journal.pone.0160890.

13. Wang, Y., Benz, F.W., Wu, Y., Wang, Q., Chen, Y., Chen, X., Li, H., Zhang, Y., Zhang, R. & Yang, J., 2016, Structural Insights into the pharmacophore of vinca domain inhibitors of microtubules, *Molecular Pharmacology,* **89**, pg. 233-242. DOI 10.1124/mol.115.100149.

14. Bai, R., Pettit, G.R. & Hamel, E., 1990, Binding of dolastatin 10 to tubulin at a distinct site for peptide antimitotic agents near the exchangeable nucleotide and vinca alkaloid sites, *Journal of Biological Chemistry,* **265**(28), pg. 17141-17149.

15. Jordan, M.A., Thrower, D. & Wilson, L., 1991, Mechanism of Inhibition of Cell Proliferation by Vinca Alkaloids, *Cancer Research,* **51**(8), pg. 2212-2222.

16. Dividing Cells Showing Chromosomes and Cell Skeleton, Nasser Rusan, National Heart, Lung, and Blood Institute, National Institutes of Health (2014).

17. Johansson, M. P., Maaheimo, H., Ekholm, F. S., 2017, New insight on the structural features of the cytotoxic auristatins MMAE and MMAF revealed by combined NMR spectroscopy and quantum chemical modelling, *Scientific Reports*, **7**, pg. 1-10. DOI 10.1038/s41598-017-15674-1.

18. Adcetris - Summary of product characteristics, Takeda Pharma A/S (2018).

19. Younes, A. & Kadin, M.E., 2003, Emerging applications of the tumor necrosis factor family of ligands and receptors in cancer therapy, *Journal of Clinical Oncology,* **21**, pg. 3526-3534. DOI 10.1200/JCO.2003.09.037.

20. Dornan, D., Bennett, F., Chen, Y., Dennis, M., Eaton, D., Elkins, K., French, D., Go, M.A.T., Jack, A., Junutula, J.R., Koeppen, H., Lau, J., McBride, J., Rawstron, A., Shi, X., Yu, N., Yu, S.-F., Yue, P., Zheng, B., Ebens, A. & Polson, A.G., 2009, Therapeutic potential of an anti-CD79b antibody-drug conjugate, anti-CD79b-vc-MMAE, for the treatment of non-Hodgkin lymphoma, *Blood,* **114**, pg. 2721-2729. DOI 10.1182/blood-2009-02-205500.

21. Nakazato, S., Nomoto, K., Kazahari, K. & Ono, M., 1998, Physical Linkage of the B29/Ig-β (CD79B) Gene to the Skeletal Muscle, Sodium-Channel, and Growth Hormone Genes in Rat and Human, *Genomics,* **48**, pg. 363-368. DOI 10.1006/geno.1997.5178.

22. Wood, W.J., Jr., Thompson, A.A., Korenberg, J., Chen, X.-N., May, W., Wall, R. & Denny, C.T., 1993, Isolation and Chromosomal Mapping of the Human Immunoglobulin-Associated B29 Gene (IGB), *Genomics,* **16**, pg. 187-192. DOI 10.1006/geno.1993.1157.

23. FDA approves polatuzumab vedotin-piiq for diffuse large B-cell lymphoma, 2019, *Case Medical Research*. DOI 10.31525/cmr-14d2e3c.

24. *International Nonproprietary Names (INN) for pharmaceutical substances. Names for radicals, groups & others: Comprehensive list,* 2015, WHO Document Production Services, Geneva, Switzerland.

25. Newman, D.J., 2019, The "utility" of highly toxic marine-sourced compounds, *Marine Drugs,* **17**. DOI 10.3390/md17060324.

26. *European Medicines Agency - Human medicines - Rare disease (orphan) designations*. Available: http://www.ema.europa.eu/ema/ (Retrieved Jun 18, 2018).

27. *Orphan incentives*. Available: https://www.ema.europa.eu/en/human-regulatory/research-development/orphan-designation/orphan-incentives (Retrieved Aug 7, 2019).

28. *EU/3/08/596*. Available: https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu308596 (Retrieved Aug 5, 2019).

29. *EU/3/18/2013*. Available: https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu3182013 (Retrieved Aug 5, 2019).

30. *EU/3/17/1925*. Available: https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu3171925 (Retrieved Aug 5, 2019).

31. *EU/3/14/1305*. Available: https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu3141305 (Retrieved Aug 5, 2019).

32. *The Collected Papers of Paul Ehrlich, Volume III Chemotherapy,* 1960, Pergamon Press.

33. Winau, F., Westphal, O. & Winau, R., 2004, Paul Ehrlich - In search of the magic bullet, *Microbes and Infection,* **6**, pg. 786-789. DOI 10.1016/j.micinf.2004.04.003.

34. Strebhardt, K. & Ullrich, A., 2008, Paul Ehrlich's magic bullet concept: 100 Years of progress, *Nature Reviews Cancer,* **8**, pg. 473-480. DOI 10.1038/nrc2394.

35. Schrama, D., Reisfeld, R.A. & Becker, J.C., 2006, Antibody targeted drugs as cancer therapeutics, *Nature Reviews Drug Discovery,* **5**, pg. 147-159. DOI 10.1038/nrd1957.

36. Lucas, A.T., Price, L.S.L., Schorzman, A.N., Storrie, M., Piscitelli, J.A., Razo, J. & Zamboni, W.C., 2018, Factors Affecting the Pharmacology of Antibody–Drug Conjugates, *Antibodies,* **7**, pg. 10. DOI 10.3390/antib7010010.

37. Kovtun, Y.V. & Goldmacher, V.S., 2007, Cell killing by antibody-drug conjugates, *Cancer Letters,* **255**, pg. 232-240. DOI 10.1016/j.canlet.2007.04.010.

38. Stack, G.D. & Walsh, J.J., 2012, Optimising the delivery of tubulin targeting agents through antibody conjugation, *Pharmaceutical Research*, **29**, pg. 2972-2984. DOI 10.1007/s11095-012-0810-9.

39. Doronina, S.O., Toki, B.E., Torgov, M.Y., Mendelsohn, B.A., Cerveny, C.G., Chace, D.F., DeBlanc, R.L., Gearing, R.P., Bovee, T.D., Siegall, C.B., Francisco, J.A., Wahl, A.F., Meyer, D.L. & Senter, P.D., 2003, Development of potent monoclonal antibody auristatin conjugates for cancer therapy, *Nature Biotechnology*, **21**, pg. 778-784. DOI 10.1038/nbt832.

40. Doronina, S.O., Mendelsohn, B.A., Bovee, T.D., Cerveny, C.G., Alley, S.C., Meyer, D.L., Oflazoglu, E., Toki, B.E., Sanderson, R.J., Zabinski, R.F., Wahl, A.F. & Senter, P.D., 2006, Enhanced activity of monomethylauristatin F through monoclonal antibody delivery: Effects of linker technology on efficacy and toxicity, *Bioconjugate Chemistry*, **17**, pg. 114-124. DOI 10.1021/bc0502917.

41. Younes, A., Yasothan, U. & Kirkpatrick, P., 2012, Brentuximab vedotin, *Nature Reviews Drug Discovery*, **11**, pg. 19-20. DOI 10.1038/nrd3629.

42. Diagram showing a microtubule and the alpha/beta-tubulin heterodimer it's being constructed from. Thomas Splettstoesser (2015).

43. Cormier, A., Marchand, M., Ravelli, R.B.G., Knossow, M. & Gigant, B., 2008, Structural insight into the inhibition of tubulin by vinca domain peptide ligands, *EMBO Reports*, **9**, pg. 1101-1106. DOI 10.1038/embor.2008.171.

44. Gigant, B., Wang, C., Ravelli, R.B.G., Roussi, F., Steinmetz, M.O., Curmi, P.A., Sobel, A. & Knossow, M., 2005, Structural basis for the regulation of tubulin by vinblastine, *Nature*, **435**, pg. 519-522. DOI 10.1038/nature03566.

45. Bai, R., Petit, G.R. & Hamel, E., 1990, Dolastatin 10, a powerful cytostatic peptide derived from a marine animal. Inhibition of tubulin polymerization mediated through the vinca alkaloid binding domain, *Biochemical Pharmacology*, **39**, pg. 1941-1949. DOI 10.1016/0006-2952(90)90613-P.

46. Pettit, G.R., Kamano, Y., Herald, C.L., Tuinman, A.A., Boettner, F.E., Kizu, H., Schmidt, J.M., Baczynskyi, L., Tomer, K.B. & Bontems, R.J., 1987, The Isolation and Structure of a Remarkable Marine Animal Antineoplastic Constituent: Dolastatin 10, *Journal of the American Chemical Society*, **109**, pg. 6883-6885. DOI 10.1021/ja00256a070.

47. *Clinical trials webpage*. Available: https://clinicaltrials.gov/ct2/home (Retrieved June 20, 2019).

48. Bai, R., Pettit, G.R. & Hamel, E., 1990, Structure-activity studies with chiral isomers and with segments of the antimitotic marine peptide dolastatin 10, *Biochemical pharmacology*, **40**, pg. 1859-1864. DOI 10.1016/0006-2952(90)90367-T.

49. Bai, R., Roach, M.C., Jayaram, S.K., Barkoczy, J., Pettit, G.R., Luduena, R.F. & Hamel, E., 1993, Differential effects of active isomers, segments, and analogs of dolastatin 10 on ligand interactions with tubulin. Correlation with cytotoxicity, *Biochemical Pharmacology,* **45**, pg. 1503-1515. DOI 10.1016/0006-2952(93)90051-W.

50. Tumor inhibiting tetrapeptide bearing modified phenethyl amides, EP0600744A1, 1994.

51. Miyazaki, K., Kobayashi, M., Natsume, T., Gondo, M., Mikami, T., Sakakibara, K. & Tsukagoshi, S., 1995, Synthesis and Antitumor Activity of Novel Dolastatin 10 Analogs, *Chemical and Pharmaceutical Bulletin,* **43**, pg. 1706-1718. DOI 10.1248/cpb.43.1706.

52. Kobayashi, M., Natsume, T., Tamaoki, S., Watanabe, J., Asano, H., Mikami, T., Miyasaka, K., Miyazaki, K., Gondo, M., Sakakibara, K. & Tsukagoshi, S., 1997, Antitumor Activity of TZT-1027, a Novel Dolastatin 10 Derivative, *Japanese Journal of Cancer Research,* **88**, pg. 316-327. DOI 10.1111/j.1349-7006.1997.tb00383.x.

53. Natsume, T., Watanabe, J., Tamaoki, S., Fujio, N., Miyasaka, K. & Kobayashi, M., 2000, Characterization of the Interaction of TZT-1027, a Potent Antitumor Agent, with Tubulin, *Cancer Science,* **91**, pg. 737-747. DOI 10.1111/j.1349-7006.2000.tb01007.x.

54. Maderna, A., Doroski, M., Subramanyam, C., Porte, A., Leverett, C.A., Vetelino, B.C., Chen, Z., Risley, H., Parris, K., Pandit, J., Varghese, A.H., Shanker, S., Song, C., Sukuru, S.C.K., Farley, K.A., Wagenaar, M.M., Shapiro, M.J., Musto, S., Lam, M.-H., Loganzo, F. & O'Donnell, C.J., 2014, Discovery of cytotoxic dolastatin 10 analogues with N-terminal modifications, *Journal of Medicinal Chemistry*, **57**, pg. 10527-10543. DOI 10.1021/jm501649k.

55. Senter, P.D. & Sievers, E.L., 2012, The discovery and development of brentuximab vedotin for use in relapsed Hodgkin lymphoma and systemic anaplastic large cell lymphoma, *Nature Biotechnology,* **30**, pg. 631-637. DOI 10.1038/nbt.2289.

56. Li, F., Emmerton, K.K., Jonas, M., Zhang, X., Miyamoto, J.B., Setter, J.R., Nicholas, N.D., Okeley, N.M., Lyon, R.P., Benjamin, D.R. & Law, C.-L., 2016, Intracellular released payload influences potency and bystander-killing effects of antibody-drug conjugates in preclinical models, *Cancer Research,* **76**, pg. 2710-2719. DOI 10.1158/0008-5472.CAN-15-1795.

57. Sokka, I.K., Ekholm, F.S. & Johansson, M.P., 2019, Increasing the Potential of the Auristatin Cancer-Drug Family by Shifting the Conformational Equilibrium, *Molecular Pharmaceutics,* **16**, pg. 3600-3608. DOI 10.1021/acs.molpharmaceut.9b00437.

58. Kung Sutherland, M.S., Sanderson, R.J., Gordon, K.A., Andreyka, J., Cerveny, C.G., Yu, C., Lewis, T.S., Meyer, D.L., Zabinski, R.F., Doronina, S.O., Senter, P.D., Law, C.-L. & Wahl, A.F., 2006, Lysosomal trafficking and cysteine protease metabolism confer target-specific cytotoxicity by peptide-linked anti-CD30-auristatin conjugates, *Journal of Biological Chemistry,* **281**, pg. 10540-10547. DOI 10.1074/jbc.M510026200.

59. Kevin J. Hamblett, Peter D. Senter, Dana F. Chace, Michael M. C. Sun, Joel Lenox, Charles G. Cerveny, Kim M. Kissler, Starr X. Bernhardt, Anastasia K. Kopcha, Roger F. Zabinski, Damon L. Meyer & Joseph A. Francisco, 2004, Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate, *Clinical Cancer Research*, **10**, pg. 7063-7070. DOI 10.1158/1078-0432.CCR-04-0789.

60. Doronina, S.O., Bovee, T.D., Meyer, D.W., Miyamoto, J.B., Anderson, M.E., Morris-Tilden, C.A. & Senter, P.D., 2008, Novel peptide linkers for highly potent antibody-auristatin conjugate, *Bioconjugate Chemistry*, **19**, pg. 1960-1963. DOI 10.1021/bc800289a.

61. Ekholm, F.S., Pynnönen, H., Vilkman, A., Pitkänen, V., Helin, J., Saarinen, J. & Satomaa, T., 2016, Introducing Glycolinkers for the Functionalization of Cytotoxic Drugs and Applications in Antibody–Drug Conjugation Chemistry, *ChemMedChem*, **11**, pg. 2501-2505. DOI 10.1002/cmdc.201600372.

62. Burns, K.E., Robinson, M.K. & Thévenin, D., 2015, Inhibition of Cancer Cell Proliferation and Breast Tumor Targeting of pHLIP-Monomethyl Auristatin E Conjugates, *Molecular Pharmaceutics,* **12**, pg. 1250-1258. DOI 10.1021/mp500779k.

63. Ekholm, F.S., Ruokonen, S.-K., Redón, M., Pitkänen, V., Vilkman, A., Saarinen, J., Helin, J., Satomaa, T. & Wiedmer, S.K., 2019, Hydrophilic monomethyl auristatin e derivatives as novel candidates for the design of antibody-drug conjugates, *Separations,* **6**. DOI 10.3390/separations6010001.

64. Cunningham, D., Parajuli, K.R., Zhang, C., Wang, G., Mei, J., Zhang, Q., Liu, S. & You, Z., 2016, Monomethyl Auristatin E Phosphate Inhibits Human Prostate Cancer Growth. *Prostate (Hoboken, NJ, United States),* **76**, pg. 1420-1430. DOI 10.1002/pros.23226.

65. Akaiwa, M., Martin, T. & Mendelsohn, B.A., 2018, Synthesis and Evaluation of Linear and Macrocyclic Dolastatin 10 Analogues Containing Pyrrolidine Ring Modifications, *ACS Omega,* **3**, pg. 5212-5221. DOI 10.1021/acsomega.8b00093.

66. Maderna, A. & Leverett, C.A., 2015, Recent advances in the development of new auristatins: Structural modifications and application in antibody drug conjugates, *Molecular Pharmaceutics,* **12**, pg. 1798-1812. DOI 10.1021/mp500762u.

67. Dugal-Tessier, J., Barnscher, S.D., Kanai, A. & Mendelsohn, B.A., 2017, Synthesis and Evaluation of Dolastatin 10 Analogues Containing Heteroatoms on the Amino Acid Side Chains, *Journal of Natural Products,* **80**, pg. 2484-2491. DOI 10.1021/acs.jnatprod.7b00359.

68. Boron, W.F. & Boulpaep, E.L., 2017, *Medical Physiology*, 3rd ed., Saunders/Elsevier.

69. Lu, D., Gillespie, W.R., Girish, S., Agarwal, P., Li, C., Hirata, J., Chu, Y.-W., Kagedal, M., Leon, L., Maiya, V. & Jin, J.Y., 2017, Time-to-event analysis of polatuzumab vedotin-

induced peripheral neuropathy to assist in the comparison of clinical dosing regimens, *CPT: Pharmacometrics and Systems Pharmacology*, **6**, pg. 401-408. DOI 10.1002/psp4.12192.

70. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe Jr., E.W., 2014, Computational methods in drug discovery, *Pharmacological Reviews*, **66**, pg. 334-395. DOI 10.1124/pr.112.007336.

71. Kitchen, D.B., Decornez, H., Furr, J.R. & Bajorath, J., 2004, Docking and scoring in virtual screening for drug discovery: Methods and applications, *Nature Reviews Drug Discovery*, **3**, pg. 935-949. DOI 10.1038/nrd1549.

72. Lionta, E., Spyrou, G., Vassilatis, D.K. & Cournia, Z., 2014, Structure-based virtual screening for drug discovery: Principles, applications and recent advances, *Current Topics in Medicinal Chemistry*, **14**, pg. 1923-1938. DOI 10.2174/1568026614666140929124445.

73. Plewczynski, D., Lazniewski, M., Augustyniak, R. & Ginalski, K., 2011, Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database, *Journal of Computational Chemistry*, **32**, pg. 742-755. DOI 10.1002/jcc.21643.

74. Hawkins, P.C.D., 2017, Conformation Generation: The State of the Art, *Journal of Chemical Information and Modeling*, **57**, pg. 1747-1756. DOI 10.1021/acs.jcim.7b00221.

75. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. & Olson, A.J., 1998, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry*, **19**, pg. 1639-1662. DOI AID-JCC10>3.0.CO;2-B.

76. Ryde, U. & Söderhjelm, P., 2016, Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods, *Chemical Reviews*, **116**, pg. 5520-5566. DOI 10.1021/acs.chemrev.5b00630.

77. Jensen, F., 2017, *Introduction to Computational Chemistry*, 3rd ed., Wiley, Chichester, UK.

78. Fock, V., 1930, Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Zeitschrift für Physik*, **61**, pg. 126-148. DOI 10.1007/BF01340294.

79. Hartree, D.R., 1928, The Wave Mechanics of an Atom with a Non-Coulomb Central Field Part I Theory and Methods, *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**, pg. 89-110. DOI 10.1017/S0305004100011919.

80. Schaefer III, H.F., 1988, A history of ab initio computational quantum chemistry: 1950-1960, *Tetrahedron Computer Methodology*, **1**, pg. 97-102. DOI 10.1016/0898-5529(88)90014-0.

81. Perdew, J.P., Ruzsinszky, A., Constantin, L.A., Sun, J. & Csonka, G.I., 2009, Some fundamental issues in ground-state density functional theory: A guide for the perplexed, *Journal of Chemical Theory and Computation*, **5**, pg. 902-908. DOI 10.1021/ct800531s.

82. Møller, C. & Plesset, M.S., 1934, Note on an approximation treatment for many-electron systems, *Physical Review,* **46**, pg. 618-622. DOI 10.1103/PhysRev.46.618.

83. Bartlett, R.J. & Musial, M., 2007, Coupled-cluster theory in quantum chemistry, *Reviews of Modern Physics,* **79**, pg. 291-352. DOI 10.1103/RevModPhys.79.291.

84. Čížek, J., 1966, On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods, *The Journal of Chemical Physics,* **45**, pg. 4256-4266. DOI 10.1063/1.1727484.

85. Hartree, D.R., 1928, The Wave Mechanics of an Atom with a Non-Coulomb Central Field Part II Some Results and Discussion, *Mathematical Proceedings of the Cambridge Philosophical Society,* **24**, pg. 111-132. DOI 10.1017/S0305004100011920.

86. Rumble, J.R., Lide, D.R. & Bruno, T.J., 2019, *CRC Handbook of Chemistry and Physics*, (Internet Version 2019) 100th ed., CRC Press / Taylor & Francis, Boca Raton, FL.

87. Pople, J.A., Binkley, J.S. & Seeger, R., 1976, Theoretical models incorporating electron correlation, *International Journal of Quantum Chemistry,* **10**, pg. 1-19. DOI 10.1002/qua.560100802.

88. Grimme, S., 2003, Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies, *Journal of Chemical Physics,* **118**, pg. 9095-9102. DOI 10.1063/1.1569242.

89. Grimme, S., Goerigk, L. & Fink, R.F., 2012, Spin-component-scaled electron correlation methods, *Wiley Interdisciplinary Reviews: Computational Molecular Science,* **2**, pg. 886-906. DOI 10.1002/wcms.1110.

90. Liakos, D.G., Sparta, M., Kesharwani, M.K., Martin, J.M.L. & Neese, F., 2015, Exploring the accuracy limits of local pair natural orbital coupled-cluster theory, *Journal of Chemical Theory and Computation,* **11**, pg. 1525-1539. DOI 10.1021/ct501129s.

91. Liakos, D.G. & Neese, F., 2015, Is It Possible to Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory, *Journal of Chemical Theory and Computation,* **11**, pg. 4054-4063. DOI 10.1021/acs.jctc.5b00359.

92. Riplinger, C. & Neese, F., 2013, An efficient and near linear scaling pair natural orbital based local coupled cluster method, *Journal of Chemical Physics,* **138**. DOI 10.1063/1.4773581.

93. Riplinger, C., Sandhoefer, B., Hansen, A. & Neese, F., 2013, Natural triple excitations in local coupled cluster calculations with pair natural orbitals, *Journal of Chemical Physics,* **139**. DOI 10.1063/1.4821834.

94. Foster, J.M. & Boys, S.F., 1960, Canonical configurational interaction procedure, *Reviews of Modern Physics,* **32**, pg. 300-302. DOI 10.1103/RevModPhys.32.300.

95. Boys, S.F., 1960, Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another, *Reviews of Modern Physics,* **32**, pg. 296-299. DOI 10.1103/RevModPhys.32.296.

96. Riplinger, C., Pinski, P., Becker, U., Valeev, E.F. & Neese, F., 2016, Sparse maps - A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory, *Journal of Chemical Physics,* **144**. DOI 10.1063/1.4939030.

97. Hohenberg, P. & Kohn, W., 1964, Inhomogeneous Electron Gas, *Physical Review,* **136**, pg. B86-B871. DOI 10.1103/PhysRev.136.B864.

98. Kohn, W., 1999, Nobel Lecture: Electronic structure of matter - wave functions and density functionals, *Reviews of Modern Physics,* **71**, pg. 1253-1266. DOI 10.1103/RevModPhys.71.1253.

99. Kohn, W. & Sham, L.J., 1965, Self-consistent equations including exchange and correlation effects, *Physical Review,* **140**, pg. A1133-A1138. DOI 10.1103/PhysRev.140.A1133.

100. Mardirossian, N. & Head-Gordon, M., 2017, Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals, *Molecular Physics,* **115**, pg. 2315-2372. DOI 10.1080/00268976.2017.1333644.

101. Kurth, S., Perdew, J.P. & Blaha, P., 1999, Molecular and solid-state tests of density functional approximations: LSD, GGAs, and Meta-GGAs, *International Journal of Quantum Chemistry,* **75**, pg. 889-909. DOI 10.1002/(SICI)1097-461X(1999)75:4/53.0.CO;2-8.

102. Cohen, A.J., Mori-Sánchez, P. & Yang, W., 2012, Challenges for density functional theory, *Chemical Reviews,* **112**, pg. 289-320. DOI 10.1021/cr200107z.

103. Grimme, S., 2006, Semiempirical hybrid density functional with perturbative second-order correlation, *Journal of Chemical Physics,* **124**. DOI 10.1063/1.2148954.

104. Staroverov, V.N., Scuseria, G.E., Tao, J. & Perdew, J.P., 2003, Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes, *Journal of Chemical Physics,* **119**, pg. 12129-12137. DOI 10.1063/1.1626543.

105. Wodrich, M.D., Jana, D.F., Von Ragué Schleyer, P. & Corminboeuf, C., 2008, Empirical corrections to density functional theory highlight the importance of nonbonded intramolecular interactions in alkanes, *Journal of Physical Chemistry A,* **112**, pg. 11495-11500. DOI 10.1021/jp806619z.

106. Grimme, S., Antony, J., Schwabe, T. & Mück-Lichtenfeld, C., 2007, Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules, *Organic and Biomolecular Chemistry,* **5**, pg. 741-758. DOI 10.1039/b615319b.

107. Grimme, S., Ehrlich, S. & Goerigk, L., 2011, Effect of the damping function in dispersion corrected density functional theory, *Journal of Computational Chemistry,* **32**, pg. 1456-1465. DOI 10.1002/jcc.21759.

108. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H., 2010, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, *Journal of Chemical Physics,* **132**. DOI 10.1063/1.3382344.

109. Becke, A.D. & Johnson, E.R., 2005, A density-functional model of the dispersion interaction, *Journal of Chemical Physics,* **123**. DOI 10.1063/1.2065267.

110. Parrish, R.M., Parker, T.M. & David Sherrill, C., 2014, Chemical assignment of symmetry-adapted perturbation theory interaction energy components: The functional-group SAPT partition, *Journal of Chemical Theory and Computation,* **10**, pg. 4417-4431. DOI 10.1021/ct500724p.

111. Parrish, R.M. & Sherrill, C.D., 2014, Spatial assignment of symmetry adapted perturbation theory interaction energy components: The atomic SAPT partition, *Journal of Chemical Physics,* **141**. DOI 10.1063/1.4889855.

112. Jeziorski, B., Moszynski, R. & Szalewicz, K., 1994, Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes, *Chemical Reviews,* **94**, pg. 1887-1930. DOI 10.1021/cr00031a008.

113. Weigend, F. & Ahlrichs, R., 2005, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Physical Chemistry Chemical Physics,* **7**, pg. 3297-3305. DOI 10.1039/b508541a.

114. Rappoport, D. & Furche, F., 2010, Property-optimized Gaussian basis sets for molecular response calculations, *Journal of Chemical Physics,* **133**. DOI 10.1063/1.3484283.

115. Contreras-Garcia, J., Johnson, E.R., Keinan, S., Chaudret, R., Piquemal, J., Beratan, D.N. & Yang, W., 2011, NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *Journal of Chemical Theory and Computation,* **7**, pg. 625-632. DOI 10.1021/ct100641a.

116. Johnson, E.R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A.J. & Yang, W., 2010, Revealing noncovalent interactions, *Journal of the American Chemical Society,* **132**, pg. 6498-6506. DOI 10.1021/ja100936w.

117. Huey, R., Morris, G.M., Olson, A.J. & Goodsell, D.S., 2007, A semiempirical free energy force field with charge-based desolvation, *Journal of Computational Chemistry,* **28**, pg. 1145-1152. DOI 10.1002/jcc.20634.

118. Morris, G.M., Ruth, H., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. & Olson, A.J., 2009, Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *Journal of Computational Chemistry*, **30**, pg. 2785-2791. DOI 10.1002/jcc.21256.

119. Cosconati, S., Forli, S., Perryman, A.L., Harris, R., Goodsell, D.S. & Olson, A.J., 2010, Virtual screening with AutoDock: Theory and practice, *Expert Opinion on Drug Discovery,* **5**, pg. 597-607. DOI 10.1517/17460441.2010.484460.

120. Lu, Y., Liu, Y., Xu, Z., Li, H., Liu, H. & Zhu, W., 2012, Halogen bonding for rational drug design and new drug discovery, *Expert Opinion on Drug Discovery*, **7**, pg. 375-383. DOI 10.1517/17460441.2012.678829.

121. Xu, Z., Yang, Z., Liu, Y., Lu, Y., Chen, K. & Zhu, W., 2014, Halogen bond: Its role beyond drug-target binding affinity for drug discovery and development, *Journal of Chemical Information and Modeling*, **54**, pg. 69-78. DOI 10.1021/ci400539q.

122. Wilcken, R., Zimmermann, M.O., Lange, A., Joerger, A.C. & Boeckler, F.M., 2013, Principles and applications of halogen bonding in medicinal chemistry and chemical biology, *Journal of Medicinal Chemistry,* **56**, pg. 1363-1388. DOI 10.1021/jm3012068.

123. Zhou, Y., Wang, J., Gu, Z., Wang, S., Zhu, W., Acenã, J.L., Soloshonok, V.A., Izawa, K. & Liu, H., 2016, Next Generation of Fluorine-Containing Pharmaceuticals, Compounds Currently in Phase II-III Clinical Trials of Major Pharmaceutical Companies: New Structural Trends and Therapeutic Areas, *Chemical Reviews,* **116**, pg. 422-518. DOI 10.1021/acs.chemrev.5b00392.

124. Gillis, E.P., Eastman, K.J., Hill, M.D., Donnelly, D.J. & Meanwell, N.A., 2015, Applications of Fluorine in Medicinal Chemistry, *Journal of Medicinal Chemistry,* **58**, pg. 8315-8359. DOI 10.1021/acs.jmedchem.5b00258.

125. Eichkorn, K., Weigend, F., Treutler, O. & Ahlrichs, R., 1997, Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials, *Theoretical Chemistry Accounts,* **97**, pg. 119-124. DOI 10.1007/s002140050244.

126. Klamt, A. & Schüürmann, G., 1993, COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *Journal of the Chemical Society, Perkin Transactions 2*, pg. 799-805. DOI 10.1039/P29930000799.

127. Sasmal, D.K., Ghosh, S., Das, A.K. & Bhattacharyya, K., 2013, Solvation dynamics of biological water in a single live cell under a confocal microscope, *Langmuir,* **29**, pg. 2289-2298. DOI 10.1021/la3043473.

128. Wang, W., Foley, K., Shan, X., Wang, S., Eaton, S., Nagaraj, V.J., Wiktor, P., Patel, U. & Tao, N., 2011, Single cells and intracellular processes studied by a plasmonic-based electrochemical impedance microscopy, *Nature Chemistry*, **3**, pg. 249-255. DOI 10.1038/nchem.961.

129. Halkier, A., Helgaker, T., Jørgensen, P., Klopper, W., Koch, H., Olsen, J. & Wilson, A.K., 1998, Basis-set convergence in correlated calculations on Ne, N 2 , and H 2 O, *Chemical Physics Letters*, **286**, pg. 243-252. DOI 10.1016/S0009-2614(98)00111-0.

130. Deglmann, P., May, K., Furche, F. & Ahlrichs, R., 2004, Nuclear second analytical derivative calculations using auxiliary basis set expansions, *Chemical Physics Letters*, **384**, pg. 103-107. DOI 10.1016/j.cplett.2003.11.080.

131. Deglmann, P., Furche, F. & Ahlrichs, R., 2002, An efficient implementation of second analytical derivatives for density functional methods, *Chemical Physics Letters*, **362**, pg. 511-518. DOI 10.1016/S0009-2614(02)01084-9.

132. Papajak, E., Zheng, J., Xu, X., Leverentz, H.R. & Truhlar, D.G., 2011, Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions, *Journal of Chemical Theory and Computation*, **7**, pg. 3027-3034. DOI 10.1021/ct200106a.

133. Gasteiger, J. & Marsili, M., 1980, Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges, *Tetrahedron*, **36**, pg. 3219-3228. DOI 10.1016/0040-4020(80)80168-2.

134. Bayly, C.I., Cieplak, P., Cornell, W.D. & Kollman, P.A., 1993, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model, *Journal of Physical Chemistry*, **97**, pg. 10269-10280. DOI 10.1021/j100142a004.

135. TURBOMOLE v.7.2, University of Karlsruhe and Forschungszentrum Karlsruhe GmbH. (2017), http://www.turbomole.com/.

136. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C., 1989, Electronic structure calculations on workstation computers: The program system TURBOMOLE, *Chemical Physics Letters*, **162**, pg. 165-169. DOI 10.1016/0009-2614(89)85118-8.

137. Neese, F., 2012, The ORCA program system, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2**, pg. 73-78. DOI 10.1002/wcms.81.

138. Valiev, M., Bylaska, E.J., Govind, N., Kowalski, K., Straatsma, T.P., Van Dam, H. J. J., Wang, D., Nieplocha, J., Apra, E., Windus, T.L. & De Jong, W.A., 2010, NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations, *Computer Physics Communications*, **181**, pg. 1477-1489. DOI 10.1016/j.cpc.2010.04.018.

139. Parrish, R.M., Burns, L.A., Smith, D.G.A., Simmonett, A.C., DePrince, A.E., Hohenstein, E.G., Bozkaya, U., Sokolov, A.Y., Di Remigio, R., Richard, R.M., Gonthier, J.F., James,

A.M., McAlexander, H.R., Kumar, A., Saitow, M., Wang, X., Pritchard, B.P., Verma, P., Schaefer, H.F., Patkowski, K., King, R.A., Valeev, E.F., Evangelista, F.A., Turney, J.M., Crawford, T.D. & Sherrill, C.D., 2017, Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability, *Journal of Chemical Theory and Computation,* **13**, pg. 3185-3197. DOI 10.1021/acs.jctc.7b00174.

140. Jmol: An Open-Source Java Viewer for Chemical Structures in 3D, http://www.jmol.org/.

141. Humphrey, W., Dalke, A. & Schulten, K., 1996, VMD: Visual Molecular Dynamics, *Journal of Molecular Graphics,* **14**, pg. 33-38. DOI 10.1016/0263-7855(96)00018-5.

142. Besora, M., Vidossich, P., Lledós, A., Ujaque, G. & Maseras, F., 2018, Calculation of Reaction Free Energies in Solution: A Comparison of Current Approaches, *Journal of Physical Chemistry A,* **122**, pg. 1392-1399. DOI 10.1021/acs.jpca.7b11580.

143. Ho, J., Klamt, A. & Coote, M.L., 2010, Comment on the correct use of continuum solvent models, *Journal of Physical Chemistry A,* **114**, pg. 13442-13444. DOI 10.1021/jp107136j.

144. Ribeiro, R.F., Marenich, A.V., Cramer, C.J. & Truhlar, D.G., 2011, Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation, *Journal of Physical Chemistry B,* **115**, pg. 14556-14562. DOI 10.1021/jp205508z.